GAUSSIAN SEMIPARAMETRIC ESTIMATION OF NON STATIONARY TIME SERIES

By Carlos Velasco

Universidad Carlos III de Madrid

First version received June 1997

Abstract. Generalizing the definition of the memory parameter d in terms of the differentiated series, we showed in Velasco (Non-stationary log-periodogram regression, Forthcoming *J. Economet.*, 1997) that it is possible to estimate consistently the memory of non-stationary processes using methods designed for stationary long-range-dependent time series. In this paper we consider the Gaussian semiparametric estimate analysed by Robinson (Gaussian semiparametric estimation of long range dependence. *Ann. Stat*. 23 (1995), 1630 61) for stationary processes. Without *a priori* knowledge about the possible non-stationarity of the observed process, we obtain that this estimate is consistent for $d \in (\ \frac{1}{2}, 1)$ and asymptotically normal for $d \in (\ \frac{1}{2}, \frac{3}{4})$ under a similar set of assumptions to those in Robinson's paper. Tapering the observations, we can estimate any degree of non-stationarity, even in the presence of deterministic polynomial trends of time. The semiparametric efficiency of this estimate for stationary sequences also extends to the non-stationary framework.

Keywords. Non-stationary time series; semiparametric inference; tapering.

1. INTRODUCTION

Statistical inference for stationary long range dependent time series is often based on semiparametric estimates that avoid parameterization of the short run behaviour. Frequently, it is assumed that the spectral density $f(\lambda)$ of the observed stationary sequence satisfies, for $0 < G < \infty$,

$$f(\lambda) \sim G\lambda^{\ \ 2d} \qquad \text{as } \lambda \to 0^+ \tag{1}$$

where $d \in (-\frac{1}{2}, \frac{1}{2})$ is the parameter that governs the degree of memory of the series. This is the interval of values of d for which the process is stationary and invertible. If $d \in (0, \frac{1}{2})$ then we say that the series exhibits long memory or long range dependence. When $d < 0$ the spectral density satisfies $f(0) = 0$ and if $d \leqslant -\frac{1}{2}$ the process is not invertible. Many non stationary time series are transformed into stationary time series after taking a sufficient number of differences. In this case it is straightforward to generalize the definition of the memory parameter d in terms of the properties of the spectral density of the stationary increments of the observed process and the unit root filter(s). Robinson (1995a) recommended an initial, possibly repeated, differentiation (integration) of the observed time series when non stationarity (non invertibility) is suspected, to obtain a value of d in the stationary and invertible interval $(\ \frac{1}{2}, \frac{1}{2})$, and then apply stationary procedures on the transformed series, adjusting the estimate with the number of differences (integrations) taken.

However, in many empirical applications values of d outside the stationary range are found when the estimates are not constrained to the stationary range $d < \frac{1}{2}$, as is the case of explicit form estimates like the log periodogram regression (e.g. Bloomfield, 1991; Agiakloglou *et al.*, 1993). In Velasco (1997a) we considered the application of the log periodogram regression estimate (see Geweke and Porter Hudak, 1983; Robinson, 1995a) to raw non stationary processes, following some previous ideas in Hassler (1992) and Hurvich and Ray (1995). The last authors considered the expectation of the periodogram at low Fourier frequencies for non stationary and non invertible fractionally integrated processes. They showed that the normalized periodogram has bounded expectation for $d \in [\frac{1}{2}, \frac{3}{2})$ but it is biased (for a function f satisfying (1)) in this case.

Robinson (1995b) found that in the stationary and invertible case an estimate of d minimizing an approximation to a Gaussian likelihood for frequencies close to the origin has better efficiency properties than rival semiparametric estimates, in the sense of having smaller asymptotic variance after proper normalization when using the same amount of sample information. Using Velasco's (1997a) results for the periodogram of non stationary time series, we address in this paper whether it is possible to extend the range of allowed values of d in this implicitly defined estimate to cover some non stationary situations and what the properties of the estimates are when the series is non stationary, including some possible efficiency gains.

Under similar conditions to those assumed by Robinson we find that the Gaussian semiparametric estimate is consistent for $d \in (\ \frac{1}{2}, 1)$ and asympto tically normal for $d < \frac{3}{4}$, with the same variance as in the stationary situation, being more efficient than the log periodogram regression estimator. If we taper the observations adequately we can estimate higher degrees of non stationarity, as was found for the log periodogram estimate in Velasco (1997a). Finally, we perform a limited numerical study of these theoretical results with simulated and real data. We give all the proofs together with some technical lemmas at the end of the paper in two appendices.

We do not discuss the non invertible case $d \leqslant \ \frac{1}{2}$ here, but this could be done using similar methods to those of Velasco (1997a) for the log periodogram estimate (see Theorems 9 and 10 in that paper).

2. ASSUMPTIONS AND DEFINITIONS

In Sections 2 and 3 we consider the original estimate analysed by Robinson (1995b) and concentrate on the interval $\ \frac{1}{2} < d < \frac{3}{2}$. When the observed time series is stationary with spectral density $f_X(\lambda)$ satisfying (1), $d < \frac{1}{2}$, we say that the process has memory d and we define the function f as

$$f(\lambda) = f_X(\lambda).$$

When $\{X_t\}$ is a non stationary process, we say that it has memory parameter $d\,(\frac{1}{2} \leqslant d < \frac{3}{2})$ if the zero mean stationary process $U_t = \Delta X_t$ has spectral density

$$f_U(\lambda) = |1 - \exp(i\lambda)|^{-2(d-1)} f^*(\lambda)$$

where $f^*(\lambda)$ is a spectral density on $[-\pi, \pi]$ which is bounded above and away from zero and is continuous at $\lambda = 0$. Thus $f_U(\lambda)$ satisfies (1) with some $-\frac{1}{2} \leqslant d_U < \frac{1}{2}$, but we do not restrict its form for frequencies away from the origin. Then we assume, following Hurvich and Ray (1995), that for any $t \geqslant 1$

$$X_t = \sum_{k=1}^{t} U_k + X_0$$

where X_0 is a random variable not depending on time t. Next, define the function $f(\lambda)$ for $d \geqslant \frac{1}{2}$:

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2} f_U(\lambda) = |1 - \exp(i\lambda)|^{-2d} f^*(\lambda) = |2\sin(\lambda/2)|^{-2d} f^*(\lambda).$$

Note that f satisfies (1), but when $2d \geqslant 1$ it is not integrable in $[-\pi, \pi]$ and is not a spectral density. We do not assume that f^* is the spectral density of a stationary and invertible autoregressive moving average (ARMA) process as would be the case if U_t followed a fractional autoregressive integrated moving average (ARIMA) model. Here f^* may have (integrable) poles or zeros at frequencies beyond the origin.

We want to give a unified theory for semiparametric estimates of $d \in (-\frac{1}{2}, 1)$, including stationary (with $f_X(0)$ equal to zero, a constant or infinity) and non stationary processes. We introduce now the following assumptions about the behaviour of the spectral densities $f_X(\lambda)$ $(d < \frac{1}{2})$ and $f_U(\lambda)$ $(d \geqslant \frac{1}{2})$ (and thus of the functions $f(\lambda)$ and $f^*(\lambda)$) at the origin.

Assumption 1. When $d \in (-\frac{1}{2}, \frac{1}{2})$ the spectral density $f_X(\lambda)$ satisfies, for $0 < G < \infty$,

$$f_X(\lambda) \sim G\lambda^{-2d} \qquad \text{as } \lambda \to 0^+$$

and when $d \in [\frac{1}{2}, \frac{3}{2})$ the spectral density $f_U(\lambda)$ satisfies

$$f_U(\lambda) \sim G\lambda^{-2(d-1)} \qquad \text{as } \lambda \to 0^+.$$

A slightly stronger version of this assumption, and the one we shall use to obtain the asymptotic normality of our estimates, is the following.

Assumption 2. When $d \in (-\frac{1}{2}, \frac{1}{2})$ the spectral density $f_X(\lambda)$ satisfies, for $0 < \beta \leqslant 2,\ 0 < G < \infty$,

$$f_X(\lambda) = G\lambda^{-2d} + \mathrm{O}(\lambda^{-2d+\beta}) \qquad \text{as } \lambda \to 0^+$$

and when $d \in [\frac{1}{2}, \frac{3}{2})$ the spectral density $f_U(\lambda)$ satisfies

$$f_U(\lambda) = G\lambda^{-2(d-1)} + \mathrm{O}(\lambda^{-2(d-1)+\beta}) \qquad \text{as } \lambda \to 0^+.$$

Under Assumption 2 we write, defining the function $g(\lambda) = G\lambda^{-2d}$, $0 < \beta \leqslant 2$,

$$\frac{f(\lambda)}{g(\lambda)} = 1 + \mathrm{O}(\lambda^{\beta}) \qquad \text{as } \lambda \to 0^+. \tag{2}$$

This is equivalent to Assumption 1 in Robinson (1995a) when f is the spectral density of X_t (stationary) and $d \in (\ \frac{1}{2}, \frac{1}{2})$. See also Remark 3.1 in Giraitis *et al.* (1995).

Also, Assumption 2 implies that $f^*(\lambda)$ is bounded above and away from zero and is continuous in an interval $(0, \varepsilon)$, $\varepsilon > 0$.

Assumption 3. In a neighbourhood $(0, \varepsilon)$ of the origin, if $d \in (\ \frac{1}{2}, \frac{1}{2})$, $f_X(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} f_X(\lambda) \right| = \mathrm{O}(\lambda^{-1-2d}) \qquad \text{as } \lambda \to 0^+$$

and if $d \geqslant \frac{1}{2}$, $f_U(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} f_U(\lambda) \right| = \mathrm{O}(\lambda^{-1-2(d-1)}) \qquad \text{as } \lambda \to 0^+.$$

Then $f(\lambda)$ has first derivative satisfying (cf. Assumption 2 of Robinson (1995a) in the stationary case $d < \frac{1}{2}$)

$$\left| \frac{d}{d\lambda} f(\lambda) \right| = \mathrm{O}(\lambda^{-1-2d}) \qquad \text{as } \lambda \to 0^+. \tag{3}$$

These assumption could have been formulated in terms of the functions f^* and/or f, since we are interested in the implications they have on the function f, (2) and (3). However, we did not find it appropriate to make assumptions directly on f or f^*, since these functions do not have an immediate and clear statistical interpretation as f_U or f_X have.

Now we make the following assumption about the series U_t when $d \geqslant \frac{1}{2}$, or for X_t when $d < \frac{1}{2}$, paralleling Robinson (1995b).

Assumption 4. We have, for $\ \frac{1}{2} < d < \frac{1}{2}$, $y_t = X_t$ or, for $\frac{1}{2} \leqslant d < 1$, $y_t = U_t$, with

$$y_t = \sum_{l\ 0}^{\infty} \alpha_l \epsilon_{t-l} \qquad \sum_{l\ 0}^{\infty} \alpha_l^2 < \infty$$

where

$$E(\epsilon_t | F_{t-1}) = 0 \qquad E(\epsilon_t^2 | F_{t-1}) = 1 \text{ almost surely (a.s.)} \qquad t = 0, \pm 1, \ldots$$

in which F_t is the σ field of events generated by ϵ_t, $s \leqslant t$, and there exists a random variable ϵ such that $E\epsilon^2 < \infty$ and, for all $\eta > 0$ and some $C > 0$, $P(|\epsilon_t| > \eta) \leqslant CP(|\epsilon| > \eta)$.

Then we obtain that, for $d \geqslant \frac{1}{2}$,

$$f(\lambda) = |1 \quad \exp(i\lambda)|^{-2} f_U(\lambda) = |1 \quad \exp(i\lambda)|^{-2} \frac{|\alpha(\lambda)|^2}{2\pi}$$

where

$$\alpha(\lambda) = \sum_{l \quad 0}^{\infty} \alpha_l \exp(il\lambda)$$

and $|\alpha(\lambda)|^2/2\pi = f_U(\lambda)$, the spectral density of U_t.

Define the discrete Fourier transform of X_t, $t = 1, \ldots, n$, $\lambda_j = 2\pi j/n$, j integer,

$$w(\lambda_j) = \frac{1}{(2\pi n)^{1/2}} \sum_{t \quad 1}^{n} X_t \exp(i\lambda_j t)$$

and when $d \geqslant \frac{1}{2}$ we obtain

$$w(\lambda_j) = \frac{1}{(2\pi n)^{1/2}} \sum_{t \quad 1}^{n} \sum_{k \quad 1}^{t} U_k \exp(i\lambda_j t)$$

so $w(\lambda_j)$ is a complex linear combination of the (non observable stationary variables U_k. The Fourier transform at any frequency λ_j, $0 < j < n$, of a non stationary sequence X_t allows the elimination of the random variable X_0, so $w(\lambda_j)$ does not depend on the values of U_k for $k < 1$. Define the periodogram of X_t as

$$I(\lambda_j) = |w(\lambda_j)|^2.$$

Because the estimate is not defined in closed form, we denote by G_0 and d_0 the true parameter values, and by G and d any admissible values. Consider the objective function (see Künsch, 1987; Robinson, 1995b)

$$Q(G, d) = \frac{1}{m} \sum_{j \quad 1}^{m} \left\{ \log(G\lambda_j^{-2d}) + \frac{I(\lambda_j)}{G\lambda_j^{-2d}} \right\}$$

and define the closed interval of admissible estimates of d_0, $\Theta = [\nabla_1, \nabla_2]$, where ∇_1 and ∇_2 are numbers such that $\frac{1}{2} < \nabla_1 < \nabla_2 < 1$. Note that we cover part of the range of values of d for which X_t is non stationary. As in Robinson (1995b) ∇_1 and ∇_2 can be chosen arbitrarily close to $\frac{1}{2}$ and 1 ($\frac{1}{2}$ in his case), respectively, or reflecting some prior knowledge on d_0. When $d_0 \in (\ \frac{1}{2}, \frac{1}{2})$ the asymptotics for $I(\lambda_j)$ are exactly the same as in Robinson's discussion, but when $d_0 \geqslant \frac{1}{2}$ we have to resort to the results of Velasco (1997a), weaker in general. Robinson used notation in terms of the parameter $H = d + \frac{1}{2}$, but we find it more natural to use the number of differences parameter d in a possibly non stationary context. We define the estimates

$$(\hat{G}, \hat{d}) = \arg \min_{0 < G < \infty, d \in \Theta} Q(G, d)$$

which always exist and also

$$\hat{d} = \arg \min_{d \in \Theta} R(d)$$

where

$$R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j \qquad \hat{G}(d) = \frac{1}{m} \sum_{1}^{m} \lambda_j^{2d} I(\lambda_j).$$

Using the discussion in Velasco (1997a), the main way of showing that Robinson's (1995b) results go through in the non stationary case ($d_0 \geqslant \frac{1}{2}$) is to analyse the asymptotic behaviour of the discrete Fourier transform of X_t for frequencies λ_j, $1 \leqslant j \leqslant m$, with $1/m + m/n \to 0$ as $n \to \infty$. Therefore, assuming the same conditions for the ϵ_k, we could repeat the steps in Robinson (1995b) to obtain the consistency and asymptotic distribution of the estimate of the parameter d for non stationary processes. However, because of a bias problem, the same results as in Robinson (1995b) can only be obtained for $d_0 < \frac{3}{4}$, consistency holding for $d_0 < 1$.

We stress the point that the discrete sum in the previous definitions *cannot* be substituted by an integral form as is considered for related estimates in a full parametric context (see Fox and Taqqu, 1986; Giraitis and Surgailis, 1990), since the properties of the periodogram for non stationary processes are only equivalent to the stationary case when evaluated at frequencies λ_j, $1 \leqslant j \leqslant n - 1$.

3. CONSISTENCY

In this section we obtain the consistency of $\hat{d}$ as defined previously for values $d_0 \in (-\frac{1}{2}, 1)$. Under Assumptions 2 and 3, the conditions on the behaviour of the function $f(\lambda)$ at the origin according to Theorem 1 in Robinson (1995b) hold now also for $d_0 \in [\frac{1}{2}, \frac{3}{2})$ (we do not need the integrability of f).

For the stationary case, the analysis of the asymptotic properties of $w(\lambda_j)$ has been done by Robinson (1995a). For the non stationary situation, $d \geqslant \frac{1}{2}$, following some ideas of Hurvich and Ray (1995) we obtain that

$$E\{I(\lambda_j)\} = \int_{-\pi}^{\pi} f(\lambda) K(\lambda - \lambda_j) d\lambda$$

where $K(\lambda) = (2\pi n)^{-1} |\sum_1^n \exp(i\lambda t)|^2$ is the Fejér kernel. From this expression it is possible to see that, when X_t is non stationary, $f(\lambda)$ plays exactly the same role as a spectral density in the asymptotics for the discrete Fourier transform at frequencies λ_j, $j \neq 0 \mod n$, and Velasco (1997a) showed that the periodogram is (asymptotically) unbiased for f if j is growing slowly with n and $d < 1$. This is stated the next theorem, which is Theorem 1 in Velasco (1997a). Definining $v(\lambda) = w(\lambda)/f^{1/2}(\lambda)$, we have the following.

THEOREM 1. *Under Assumptions 1 and 3, $d \in [\frac{1}{2}, 1)$, for any sequences of positive integers $j = j(n)$ and $k = k(n)$ such that $1 \leqslant k < j$ and $j/n \to 0$ as $n \to \infty$, defining*

$$\delta_{k,j} = (jk)^{d-1} \log(j+1)$$

(a) $E\{v(\lambda_j)\overline{v}(\lambda_j)\} = 1 + \mathrm{O}(\delta_{j,j})$;
(b) $E[v(\lambda_j)v(\lambda_j)\} = \mathrm{O}(\delta_{j,j})$;
(c) $E\{v(\lambda_j)\overline{v}(\lambda_k)\} = \mathrm{O}(k^{-1} \log j + \delta_{k,j})$;
(d) $E\{v(\lambda_j)v(\lambda_k)\} = \mathrm{O}(k^{-1} \log j + \delta_{k,j})$.

The next two results hold in a similar way for the log periodogram estimate of d for non stationary Gaussian time series. Here we do not need to assume Gaussianity in any form. First we show that consistency of $\hat{d}$ when $d < 1$.

THEOREM 2. *Under Assumptions 1 ($d_0 \in (\ \frac{1}{2}, 1)$), 3, 4 and*

$$\frac{1}{m} + \frac{m}{n} \to 0 \qquad \text{as } n \to \infty$$

we obtain $\hat{d} \to_{\mathrm{P}} d$.

4. ASYMPTOTIC NORMALITY

For values of $d_0 \geqslant 1$ the periodogram at frequencies λ_j is not unbiased for the function f and j increases, and therefore $\hat{d}$ cannot be consistent. Unlike for stationary processes, we can only obtain the asymptotic distribution for $\hat{d}$ in the non stationary case for a smaller range of values of d_0 ($d_0 < \frac{3}{4}$) than the interval where the estimate is consistent, $d_0 < 1$. This is due to the fact that the properties of the periodogram depend on convolutions of the function $f(\lambda)$, which deteriorate rapidly as f becomes more 'non integrable', i.e. as d_0 increases (see Theorem 1 above and Theorem 1 in Velasco (1997a), and the subsequent discussion).

We introduce two new assumptions that will be needed in the proofs.

ASSUMPTION 5. In a neighbourhood $(0, \varepsilon)$ of the origin, $\alpha(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} \alpha(\lambda) \right| = \mathrm{O}\left\{ \frac{|\alpha(\lambda)|}{\lambda} \right\} \qquad \text{as } \lambda \to 0^+.$$

Clearly Assumption 5 implies Assumption 3, since $f(\lambda) = |\alpha(\lambda)|^2/2\pi$ when $\frac{1}{2} < d_0 < \frac{1}{2}$ and $f(\lambda) = \{2\sin(\lambda/2)\}^{-2}|\alpha(\lambda)|^2/2\pi$ when $d_0 \geqslant \frac{1}{2}$.

Assumption 6. Assumption 4 holds and also

$$E(\epsilon_t^3|F_{t-1}) = \mu_3 \text{ a.s.} \qquad E(\epsilon_t^4|F_{t-1}) = \mu_4 \qquad t = 0, \pm 1, \ldots$$

for finite constants μ_3 and μ_4.

Theorem 3. *Under Assumptions 2, 5 and 6, with* $d_0 \in (\ \frac{1}{2}, \frac{3}{4})$, *and*

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \qquad \text{as } n \to \infty \tag{4}$$

we obtain

$$m^{1/2}(\hat{d} \quad d_0) \to {}_{\mathrm{D}} N(0, \tfrac{1}{4}).$$

This theorem coincides, not surprisingly, with the results of Velasco (1997a) for the log periodogram regression estimate of non stationary time series with Gaussian increments. Beyond these values of d, the slow convergence of the expectation of the periodogram to the function f leads to a slower convergence of the estimates of d. In Velasco (1997a) this problem was overcome for the log periodogram estimate using the bias reduction technique of tapering, as suggested by Hurvich and Ray (1995). We do not pursue this approach here, but the corresponding theory is similar to that obtained in the next section for general non stationary processes and tapering schemes.

Another important point is that the efficiency property of this Gaussian estimate with respect to other comparable semiparametric estimates observed by Robinson (1995b) for stationary processes holds as well for non stationary processes when the same number of periodogram ordinates, m, is used. Further, the asymptotic distribution of $\hat{d}$ does not depend on any unknown constants, not even d_0, beyond the definition of the suitable range of valid values for the theorem, which is only limited by $d < \frac{3}{4}$.

5. GENERAL NON-STATIONARY TIME SERIES

In this section we consider the estimation of the memory parameter for general non stationary time series which after a finite number of differentiations are stationary. In general, a (possibly non stationary) process $\{X_t\}$ has memory parameter $d > \ \frac{1}{2}$ if the process $\Delta^s X_t = U_t^{(s)}$, $s = \lfloor d + \frac{1}{2} \rfloor$, is stationary with mean μ, possibly different from zero, and spectral density $f_{U^{(s)}}(\lambda)$ behaving as $G\lambda^{-2(d-s)}$, $\ \frac{1}{2} \leqslant d \quad s < \frac{1}{2}$, around the origin for some positive constant G. Robinson (1995b) considered the case $s = 0$ and in Section 2 we considered the case $s = 1$, $d < 1$, $\mu = 0$.

Define the function

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2s} f_{U^{(s)}}(\lambda) = |2\sin(\lambda/2)|^{-2d} f^*(\lambda)$$

in terms of the spectral density of the stationary sequence $U_t^{(s)}$ or the function $f^*(\lambda)$. Following the discussion in Velasco (1997a), we can write for random variables $R^{(r)}$, $r = 1, \ldots, s$, which do not depend on time t

$$\begin{aligned}
X_t &= R^{(1)} + \sum_{j_1=1}^{t} U_{j_1}^{(1)} \\
&= R^{(1)} + \sum_{j_1}^{t} \left(R^{(2)} + \sum_{j_2}^{j_1} U_{j_2}^{(2)} \right) \\
&= R^{(1)} + tR^{(2)} + \sum_{j_1}^{t} \sum_{j_2}^{j_1} \left(R^{(3)} + \sum_{j_3}^{j_2} U_{j_3}^{(3)} \right) \\
&= R^{(1)} + tR^{(2)} + \frac{1}{2}(t + t^2)R^{(3)} + \sum_{j_1}^{t} \sum_{j_2}^{j_1} \sum_{j_3}^{j_2} U_{j_3}^{(3)} \\
&= \sum_{r=1}^{s} R^{(r)} p^{(r)}(t) + \mu p_\mu(t) + \sum_{j_1}^{t} \sum_{j_2}^{j_1} \cdots \sum_{j_s}^{j_{s-1}} U_{j_s}^{(*)}
\end{aligned}$$

where $p^{(r)}(t)$ are polynomials in t of order $r - 1$, $p_\mu(t)$ is a polynomial of order s and $U_t^{(*)} = U_t^{(s)} - \mu$ has zero mean and the same spectral density as $U_t^{(s)}$. These two polynomials can be regarded as the initial conditions of the observed non stationary sequence and as a deterministic trend, respectively. In Velasco (1997a) we proposed using, instead of the original series, a tapered version with a weight sequence $\{h_t\}_{t=1}^{n}$, symmetric around $\lfloor n/2 \rfloor$, such that $\max_t h_t = 1$. Hurvich and Ray (1995) used the cosine bell to analyse the expectation of the periodogram when $d < 1.5$. Other authors also (Zhurbenko, 1979; Robinson, 1986; Dahlhaus, 1988) have shown that tapering allows inference in the presence of non stationary distortions in the observed stationary time series.

We consider now the discrete Fourier transform of the tapered series $h_t X_t$:

$$\begin{aligned}
w^{\mathrm{T}}(\lambda_j) &= \frac{1}{(2\pi \sum h_t^2)^{1/2}} \sum_{t=1}^{n} h_t X_t \exp(i\lambda_j t) \\
&= \frac{1}{(2\pi \sum h_t^2)^{1/2}} \sum_{t=1}^{n} h_t \left\{ \sum_{r=1}^{s} R^{(r)} p^{(r)}(t) + \mu p_\mu(t) \right\} \exp(i\lambda_j t) \qquad (5) \\
&\quad + \frac{1}{(2\pi \sum h_t^2)^{1/2}} \sum_{t=1}^{n} h_t \sum_{j_1}^{t} \sum_{j_2}^{j_1} \cdots \sum_{j_s}^{j_{s-1}} U_{j_s}^{(*)} \exp(i\lambda_j t). \qquad (6)
\end{aligned}$$

The term (6) reflects the accumulation of information in the non stationary time series X_t, starting from $t = 1$, but the term (5) is a nuisance component of the discrete Fourier transform which comprises the information in $\{X_t\}_1^n$ from the past. To make inferences about d we make this expression (5) equal to zero for certain frequencies λ_j, using specific orthogonality properties of the weights h_t, i.e.

$$\sum_{t=1}^{n} h_t(1 + t + t^2 + \cdots + t^s)\exp(i\lambda_j t) = 0. \tag{7}$$

Observe that in the case $s = 1$ we have only required that $\sum_{t=1}^{n} h_t \exp(i\lambda_j t) = 0$, because we were assuming $\mu = 0$ to eliminate the influence of the polynomial $p^{(1)}(t) = 1$ of order 0 (a constant with respect to t). The raw Fourier transform satisfies condition (7) with $s = 0$ (but not any of higher order). In other words, without tapering we can consider $d < 1$ but always without drift.

Defining the equivalent to the Dirichlet kernel in the tapered case

$$D_p^{\mathrm{T}}(\lambda) = \sum_{t=1}^{n} h_t \exp(it\lambda)$$

we say that a sequence of data tapers $\{h_t\}_1^n$ is of order $p = 1, 2, \ldots$ if the following two conditions are satisfied.

(a) For $N = n/p$ (which we assume integer),

$$D_p^{\mathrm{T}}(\lambda) = \frac{a(\lambda)}{n^{p-1}} \left\{ \frac{\sin(n\lambda/2p)}{\sin(\lambda/2)} \right\}^p$$

where $a(\lambda)$ is a complex function, whose modulus is bounded and bounded away from zero, with $p-1$ derivatives, all bounded in modulus as n increases for $\lambda \in [-\pi, \pi]$.

(b) For one function $b = b(n)$, $0 < b < \infty$, $\forall n > 0$,

$$\sum_{t=1}^{n} h_t^2 = bn.$$

Then, it is immediate that

$$|D_p^{\mathrm{T}}(\lambda)| \leqslant \text{constant} \times \min(n, n^{1-p}|\lambda|^{-p})$$

and, with the equivalent to the Fejér kernel, $K_p^{\mathrm{T}}(\lambda) = (2\pi \sum h_t^2)^{-1} |D_p^{\mathrm{T}}(\lambda)|^2$,

$$|K_p^{\mathrm{T}}(\lambda)| \leqslant \text{constant} \times \min(n, n^{1-2p}|\lambda|^{-2p}).$$

Also we have that $D_p^{\mathrm{T}}(\lambda)$ has zeros of order p at $\lambda = \lambda_{jp}$ and that thanks to

$$\left. \frac{d^q}{(d\lambda)^q} D_p^{\mathrm{T}}(\lambda) \right|_{\lambda=\lambda_{jp}} = 0 \qquad 0 < j < N$$

$q \leqslant p-1$, condition (7) is satisfied for $s \leqslant p-1$.

If condition (7) holds, deterministic time trends up to order s can be removed in the calculation of $w^{\mathrm{T}}(\lambda_j)$ without the need to estimate them by any means. The cosine bell taper is of order 1, so its utilization is only justified in the case $d < 1.5$ with $\mu = 0$, as was shown by Velasco (1997a) for the log periodogram semiparametric estimate. Here we do not consider this tapering scheme explicitly but, given the asymptotic behaviour of tails of the kernel K_p^{T} in this case, the conclusions are equivalent to those with $p = 3$ and for $d < 1.5$.

Two examples of data tapers satisfying the above conditions are the Parzen and Zhurbenko Kolmogorov proposals (see also Alekseev (1996) for further examples and discussion). For sample size $n = 4N$, N integer, the weights given by the Parzen window

$$h_t^{\mathrm{P}} = \begin{cases} 1 - 6[\{(2t - n)/n\}^2 - |(2t - n)/n|^3] & 1 \leqslant t \leqslant N \text{ or } 3N \leqslant t \leqslant 4N \\ 2\{1 - |(2t - n)/n|\}^3 & N < t < 3N \end{cases}$$

satisfy (7) for $j = 4, 8, \ldots, n - 4$ and $s = 3$. We can obtain (see for example Percival and Walden, 1993)

$$D^{\mathrm{P}}(\lambda) = \frac{32}{n^3}\left\{3 - 2\sin^2\left(\frac{\lambda}{2}\right)\right\}\left\{\frac{\sin(n\lambda/8)}{\sin(\lambda/2)}\right\}^4 \exp\left(\frac{in\lambda}{2}\right)$$

and $\sum_{t=1}^{n}(h_t^{\mathrm{P}})^2 \sim \text{constant} \times n$. Zhurbenko (1979) used the data weights $\{h_t^{\mathrm{Z}}\}$ suggested by Kolmogorov,

$$h_t^{\mathrm{Z}} = \rho(p, N)\left\{\frac{p(N^2 - 1)}{12\pi}\right\}^{1/4} N^{-p} c_{p,N}(t)$$

where the coefficients $c_{p,N}(t)$ are given by

$$\sum_{t=0}^{p(N-1)} z^t c_{p,N}(t+1) = (1 + z + \cdots + z^{N-1})^p = \left(\frac{1 - z^N}{1 - z}\right)^p.$$

Then, it follows that

$$(2\pi \sum h_t^2)^{1/2} D^{\mathrm{Z}}(\lambda) = \rho\left\{\frac{p(N^2 - 1)}{12\pi}\right\}^{1/4}\left[\frac{1 - \exp(iN\lambda)}{N\{1 - \exp(i\lambda)\}}\right]^p$$

and hence

$$K^{\mathrm{Z}}(\lambda) = \rho^2\left\{\frac{p(N^2 - 1)}{12\pi}\right\}^{1/2}\left\{\frac{\sin^2(n\lambda/2p)}{N^2\sin^2(\lambda/2)}\right\}^p$$

where ρ is defined adequately to make K^{Z} integrate to 1 and it can be seen to be very close to 1 for p and N big enough (see Zhurbenko, 1979). Therefore, this class of taper weights for $p = 1, 2, \ldots$, fixed in the asymptotics, and $n = pN$ satisfies condition (7) with $s \leqslant p - 1$ at frequencies λ_{jp}, $0 < j < N$.

6. TAPERED ESTIMATES

In this section we obtain the consistency and asymptotic distribution of a modified version of $\hat{d}$ when we use the previous data tapers for values $d_0 > \ \frac{1}{2}$. We introduce now the following assumptions about the behaviour of the spectral density $f_{U(s)}(\lambda)$ (and thus of the functions $f(\lambda)$ and $f^*(\lambda)$) at the origin.

ASSUMPTION 7. The spectral density $f_{U(s)}(\lambda)$, $s = \lfloor d + \frac{1}{2} \rfloor$, satisfies, for some constant $0 < G < \infty$,

$$f_{U(s)}(\lambda) \sim G\lambda^{-2(d-s)} \qquad \text{as } \lambda \to 0^+. \tag{8}$$

A slightly stronger version of Assumption 2 is the following condition, where we give more information about the behaviour of the spectral density $f_{U(s)}(\lambda)$ at the origin. This extra information will be used to reduce the bias of the tapered periodogram for f as was done in Velasco (1997b) in a related context (see also Assumption 3 in Robinson (1994b)).

ASSUMPTION 8. When $d \in (\ \frac{1}{2}, \frac{1}{2})$, the spectral density $f_{U(s)}(\lambda)$ satisfies, for numbers $0 < \beta \leqslant 2$, $0 < G$, $E_\beta < \infty$.

$$f_{U(s)}(\lambda) = G\lambda^{-2(d-s)} + E_\beta \lambda^{-2(d-s)+\beta} + \mathrm{o}(\lambda^{-2(d-s)+\beta}) \qquad \text{as } \lambda \to 0^+.$$

As before, Assumption 8 implies that $f^*(\lambda)$ is bounded above and away from zero and is continuous in an interval $(0, \varepsilon)$, $\varepsilon > 0$.

We will need also the equivalent to Assumption 3.

ASSUMPTION 9. In a neighbourhood $(0, \varepsilon)$ of the origin, if $d \in (\ \frac{1}{2}, \frac{1}{2})$, $f_{U(s)}(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} f_{U(s)}(\lambda) \right| = \mathrm{O}(\lambda^{-1-2(d-s)}) \qquad \text{as } \lambda \to 0^+.$$

Then $f(\lambda)$ has first derivative satisfying (cf. Assumption 2 of Robinson (1995a) in the stationary case $d < \frac{1}{2}$),

$$\left| \frac{d}{d\lambda} f(\lambda) \right| = \mathrm{O}(\lambda^{-1-2d}) \qquad \text{as } \lambda \to 0^+.$$

Now we make the following assumption about the series $U_t^{(s)}$, equivalent to Assumption 6.

ASSUMPTION 10. We have

$$U_t^{(*)} = \sum_{l\ 0}^{\infty} \alpha_l \epsilon_{t-l} \qquad \sum_{l\ 0}^{\infty} \alpha_l^2 < \infty$$

where the ϵ_t satisfy the conditions of Assumptions 4 and 6.

Then we obtain for any $d > \ \frac{1}{2}$ that

$$f(\lambda) = |1 \quad \exp(i\lambda)|^{-2s} f_{U(s)}(\lambda) = |1 \quad \exp(i\lambda)|^{-2s} \frac{|\alpha(\lambda)|^2}{2\pi}.$$

Defining the (tapered) periodogram of X_t as

$$I_p^{\mathrm{T}}(\lambda_j) = |w_p^{\mathrm{T}}(\lambda_j)|^2$$

we consider now the objective function

$$Q_p(G, d) = \frac{p}{m} \sum_j^m \left\{ \log(G\lambda_j^{-2d}) + \frac{I_p^{\mathrm{T}}(\lambda_j)}{G\lambda_j^{-2d}} \right\}$$

where all the summations run for $j = p, 2p, \ldots, m$, assuming m/p integer, unless otherwise stated. Define the closed interval of admissible estimates of d_0, $\Theta = [\nabla_1, \nabla_2]$, where ∇_1 and ∇_2 are numbers such that $\ \frac{1}{2} < \nabla_1 < \nabla_2 < d^*$, and $p \geqslant \lfloor d^* + \frac{1}{2} \rfloor + 1$. This last condition is equivalent to $d^* < p + \frac{1}{2}$, where d^* is the maximum value of d we can estimate with tapers of order p. Note that we can cover part of the range of values of d for which X_t is non stationary. As in Robinson (1995b), ∇_1 and ∇_2 can be chosen arbitrarily close to $\ \frac{1}{2}$ and to a maximum value of d, d^*, restricted only by the order p of the taper weights used, or reflecting some prior knowledge on d_0. When $\mu = 0$ it is enough with $d^* < p$.

We define the estimates

$$(\hat{G}_p, \hat{d}_p) = \arg \min_{0 < G < \infty, d \in \Theta} Q_p(G, d)$$

which always exist and also

$$\hat{d}_p = \arg \min_{d \in \Theta} R_p(d)$$

where

$$R_p(d) = \log \hat{G}_p(d) \quad 2d \frac{p}{m} \sum_j^m \log \lambda_j, \qquad \hat{G}_p(d) = \frac{p}{m} \sum_j^m \lambda_j^{2d} I_p^{\mathrm{T}}(\lambda_j).$$

The discrete sums in the previous definitions include only frequencies λ_j, $j = p, 2p, \ldots, m$, since the properties of the periodogram for non stationary processes are only equivalent to the stationary case when evaluated at these frequencies.

When X_t is non stationary, $f(\lambda)$ plays exactly the same role as a spectral density in the asymptotics for the discrete Fourier transform at frequencies λ_j, $j \neq 0 \bmod n$, and Velasco (1997a) showed that the periodogram is (asympto tically) unbiased for f if j is growing slowly with n and p is chosen adequately. This is stated in the next theorem, which is essentially Theorem 6 in Velasco (1997a). Note that the non tapered periodogram is an estimate with

$p = 1$. Defining now $v_p^{\mathrm{T}}(\lambda) = w^{\mathrm{T}}(\lambda)/(G^{1/2}\lambda^{-d})$, for a taper of order p, we have the following.

THEOREM 4 ($p \geqslant 2$). *Under Assumptions 8 and 9* ($d > \ \ \frac{1}{2}$, $0 < \beta \leqslant 2$) *for* $f_{U^{(s)}}$, *a data taper of order* $p = 2, 3, \ldots$, *with* $p \geqslant s + 1$ *(or just* $p > d$ *if* $\mu = 0$*), for any sequences of positive integers* $k = k(n)$ *and* $j = j(n)$, $1 \leqslant k < j$ *and* $\eta = j \quad k$, *such that* $j/n \to 0$, *defining*

$$\gamma_{j,k} \equiv (jk)^{d-p} \log(j+1)$$

we get

(a) $E\{v_p^{\mathrm{T}}(\lambda_{jp})\overline{v_p^{\mathrm{T}}}(\lambda_{jp})\} = 1 + \mathrm{O}\{\min(j^{-\beta}, j^{-1}) + (j/n)^{\beta} + \gamma_{j,j}\}$;
(b) $E[v_p^{\mathrm{T}}(\lambda_{jp})v_p^{\mathrm{T}}(\lambda_{jp})\} = \mathrm{O}(j^{-p} + \gamma_{j,j})$;
(c) $E\{v_p^{\mathrm{T}}(\lambda_{jp})\overline{v_p^{\mathrm{T}}}(\lambda_{kp})\} = \mathrm{O}(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\log n + \eta^{-p} + \gamma_{k,j})$;
(d) $E\{v_p^{\mathrm{T}}(\lambda_{jp})v_p^{\mathrm{T}}(\lambda_{kp})\} = \mathrm{O}(k^{-1}\eta^{1-p} + k^{-1}\eta^{-p}\log n + \eta^{-p} + \gamma_{k,j})$.

Then we obtain the consistency of $\hat{d}_p$ in the following theorem. Note that we only require Assumption 7 for this result, not Assumption 8, which will be used to derive the asymptotic distribution of $\hat{d}$ in the next section and was used in the previous theorem because we normalized the discrete Fourier transform by $(G\lambda^{-2d})^{1/2}$ and not by $\{f(\lambda)\}^{1/2}$.

THEOREM 5. *Under Assumptions 7, 9 and 10, with* $\nabla_1 > \ \ \frac{1}{2}$ *and* $p \geqslant \lfloor \nabla_2 + \frac{1}{2} \rfloor + 1$ *such that* $d_0 \in [\nabla_1, \nabla_2]$, $p = 2, 3, \ldots$, *and*

$$\frac{1}{m} + \frac{m}{n} \to 0 \qquad \text{as } n \to \infty$$

we obtain $\hat{d}_p \to_{\mathrm{P}} d_0$.

If we assume $\mu = 0$ then we only need in fact $p > \nabla_2$ if there are only deterministic trends in X_t up to order $p \quad 1$. We do not consider here the case $p = 1$ because this is equivalent to the non tapered situation, with $\nabla_2 < 1$ (and $\mu = 0$ necessarily). With respect to Theorem 2, the only extra condition we have used is the fourth moment of the innovations ϵ_t in Assumption 10.
Then we obtain the asymptotic normality of $\hat{d}_p$.

THEOREM 6. *Under Assumptions 5, 8* ($\beta > 1$, $\nabla_1 > \ \ \frac{1}{2}$ *and* $p \geqslant \lfloor \nabla_2 + \frac{1}{2} \rfloor + 1$ *such that* $d_0 \in [\nabla_1, \nabla_2]$, $p = 2, 3, \ldots$*), 10 and*

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \qquad \text{as } n \to \infty \tag{9}$$

we obtain

$$m^{1/2}(\hat{d} \quad d_0) \to_{\mathrm{D}} \mathrm{N}(0, \tfrac{1}{4}p\Phi)$$

where

$$\Phi = \lim_{n\to\infty}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{k\;\;0,p,2p,\ldots}^{n-p} \left\{\sum_{1}^{n} h_t^2 \cos(t\lambda_k)\right\}^2. \tag{10}$$

This theorem is equivalent to the results of Velasco (1997a) for the log periodogram regression estimate of non stationary time series with Gaussian increments. There, we changed the definition of the estimate to adapt the proofs of Robinson (1995a), but here, even with the correlation between the tapered periodogram ordinates, we do not need to modify the definition of the estimate. However, the variance of the estimate is increased slightly by a factor of Φ (generally bigger than 1) because of this correlation of the tapered periodogram, owing to the lack of orthogonality of the taper weights. This Φ takes the values 1.05000, 1.00354 and 1.00086 for the Zhurbenko kernels with $p = 2$, 3, 4 respectively, implying increments of the variance of 5%, 0.35% and 0.09% for each of the data tapers (apart from the factor p in the variance of the estimate). When $\mu = 0$, the theorem is valid with just $p > \nabla_2$. If we consider the full cosine window taper $h_t = \frac{1}{2}\{1 \quad \cos(2\pi t/n)\}$, then if we regard this taper as of order $p = 3$, with the same definitions as before, $\mu = 0$ and $d < \frac{3}{2}$, Theorem 6 holds with $\Phi = 1$, but if we use all the Fourier frequencies from λ_2 to λ_m (i.e. without spacing), then $\Phi = 35/18$ (see the discussion in Velasco (1997a, 1997b)). Note also that if we take in (10) the sum across all frequencies, we obtain with Parseval's identity

$$\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{k\;\;0,1,2,\ldots}^{n-1} \left\{\sum_{1}^{n} h_t^2 \cos(t\lambda_k)\right\}^2 = n\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{t\;\;1}^{n} h_t^4$$

where the right hand side is the usual tapering variance adjustment (cf. for example Dahlhaus, 1985, expression (3)).

The increased smoothness of the function $f(\lambda)$, $\beta > 1$, is used in conjunction with the tapering to approximate the periodogram of the observed time series by that of the innovations (see the proof of Theorem 6 in Velasco (1997a) and Theorem 2 in Velasco (1997b)). Here we cannot resort to the second moments of the tapered periodogram as was done in the non tapered case, since the correlation problem just pointed out impedes further improvement of the approximations.

7. EMPIRICAL WORK

The aim of the first simulation exercise was to address the previous properties of $\hat{d}$, especially in comparison with the log periodogram regression estimate

$$\tilde{d} = \quad \frac{1}{2}\frac{\sum_{j\;\;1}^{m} \log I(\lambda_j)\{\log j \quad (1/m)\sum_{l\;\;1}^{m} \log l\}}{\sum_{j\;\;1}^{m} \log j\{\log j \quad (1/m)\sum_{l\;\;1}^{m} \log l\}}.$$

To this end we stimulated 1000 Gaussian fractional ARIMA(0, d, 0) for each value of d in 0.45(0.1)1.25, $n = 256$, and we chose a relatively small value for m, 32. We did not perform any trimming in the definition of $\tilde{d}$. The series were simulated with the S Plus function **arima.frac.diff** and the minimum of the objective function was found with the **nlmin** command. In the search for the minimum we used as initial values for d and G those obtained with the log periodogram regression, and we did not restrict the range of possible values for d. This procedure gave no problems for any value of d, indicating a relatively well behaved objective function, even for values of $d > 1$.

The box plots for the estimates are given in Figure 1, only up to $d = 1.05$. The main features of the plots are the invariance of the distributions of $\hat{d}$ and $\tilde{d}$ to the actual value of d_0 and the efficiency and smaller bias of the Gaussian estimate with respect to the log periodogram across all d_0. For $d_0 = 1.05$ ($d_0 \geqslant 1$) neither of the two estimates is consistent and this fact is reflected by the negative bias for both, in the opposite direction of the biases when $d_0 < 1$.

The basic statistics summary is contained in Table I, including the bias of the estimates, the standard deviation, the expected standard deviation from the corresponding central limit theorems and the mean square error across replications. Note that for $d_0 \geqslant \frac{3}{4}$, Theorem 3 does not hold.

In the second simulation we considered the estimation of values $d \geqslant 1$. The only modification with respect to the previous exercise was that now the series were of length $n = 512$ and $m = 100$. The values of d_0 considered were 0.95 and 1.8, one close to the borderline of the asymptotics presented in this paper for this estimate and the other well outside. The results for $\hat{d}$ and $\tilde{d}$ are given in Figure 2. In the top row of graphics we give the box plots and in the bottom row non parametric smoothed estimates of the simulated probability density of the estimates of d. The two leftmost columns of plots, for $d = 0.95$, 1.9, indicate that the two semiparametric estimates considered work relatively well

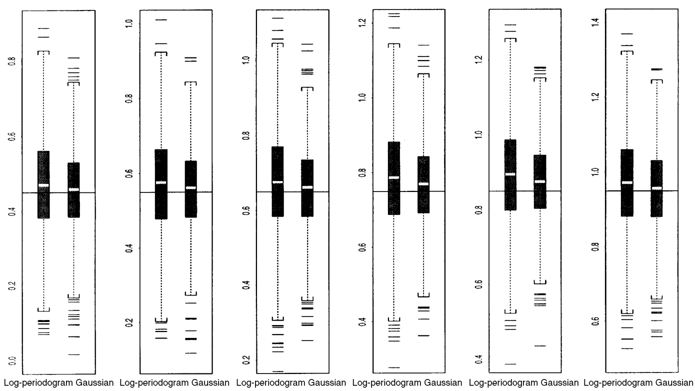

FIGURE 1. Gaussian semiparametric and log-periodogram estimates, Gaussian ARFIMA(0, d, 0), $n = 250$, $m = 32$, 1000 replications.

TABLE I

	Gaussian estimate				Log-periodogram estimate			
d_0	Bias	S.d.	Th. s.d.	MSE	Bias	S.d.	Th. s.d.	MSE
0.45	0.0041	0.1151	0.0884	0.0132	0.0179	0.1383	0.1134	0.0194
0.55	0.0050	0.1115	0.0884	0.0124	0.0186	0.1330	0.1134	0.0180
0.65	0.0089	0.1155	0.0884	0.0134	0.0259	0.1446	0.1134	0.0215
0.75	0.0164	0.1168		0.0139	0.0324	0.1439		0.0217
0.85	0.0213	0.1116		0.0129	0.0398	0.1399		0.0211
0.95	0.0026	0.1108		0.0123	0.0160	0.1342		0.0182
1.05	0.0309	0.1004		0.0110	0.0286	0.1240		0.0161
1.15	0.0837	0.0969		0.0164	0.0867	0.1207		0.0221
1.25	0.1638	0.1043		0.0377	0.1776	0.1251		0.0472

Notes: S.d., standard deviation; Th. s.d., theoretical standard deviation; MSE, mean square error.

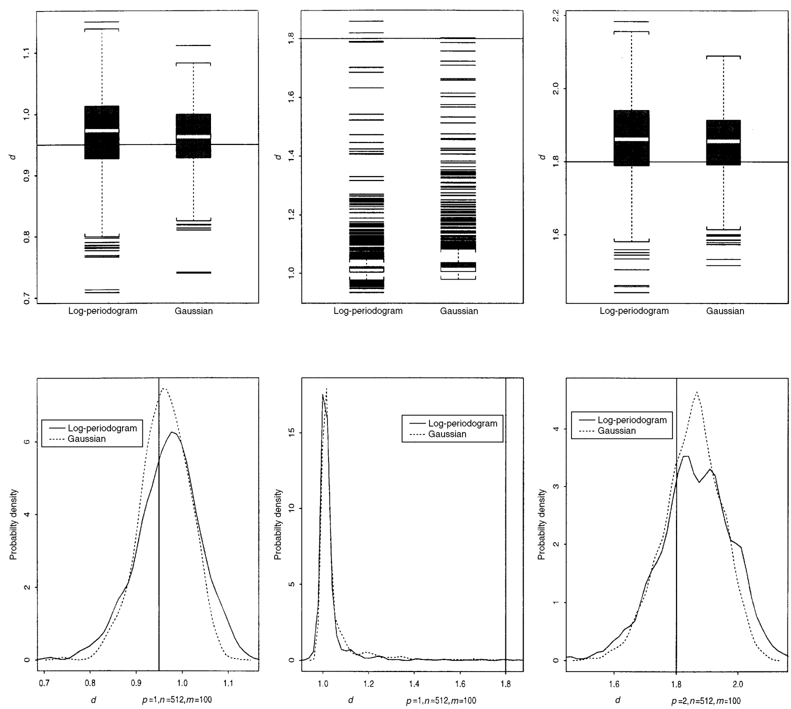

FIGURE 2. Gaussian semiparametric and log-periodogram estimates, $n = 512$, $m = 100$, Gaussian ARFIMA(0, d, 0), 1000 replications.

for values close to 1, but not for more non stationary time series, for which the estimates converge extremely quickly to values close to 1, except for a long tail towards the right value. The plots on the right are for the same estimates but we used a tapered periodogram with the triangular Barlett window taper (equivalent to Zhurbenko tapers with $p = 2$) and we defined our estimates for frequencies $\lambda_2, \lambda_4, \ldots, \lambda_m$, assuming m is even. In this case also it seems that the Gaussian estimate is more efficient than the log periodogram regression.

Now we consider a simple application with real data. Different parameter izations have been proposed in the literature to explain the persistence in the volatility of the returns found in many financial data sets. Robinson (1991) introduced a long memory generalized autoregressive conditional heteroscedas ticity (ARCH) model which was used by Baillie *et al.* (1996) and Bollerslev and Mikkelsen (1996) to define the fractionally integrated generalized ARCH class

$$\phi(L)(1 \quad L)^d x_t^2 = \omega + b(L)\nu_t$$

where all the roots of the polynomials ϕ and b in the lag operator L lie outside the unit circle and $\nu_t = x_t^2 \quad \tau_t^2$ are martingale differences, $E(\nu_t|\mathscr{F}_{t-1}) = 0$, $\tau_t^2 = \mathrm{var}(x_t|\mathscr{F}_{t-1})$ a.s. and $\mathscr{F}_t$ is the σ field of events generated by $\{x_s\colon s \leqslant t\}$.

These models allow persistence or long memory in the squares x_t^2 of martingale difference levels x_t when $d > 0$ and are basically equivalent to the fractional ARIMA models for means, but in the variance, generalizing for any $0 \leqslant d \leqslant 1$ the fully integrated GARCH model, equivalent to a unit root in the mean. Although our asymptotic theory for semiparametric estimation is not readily applicable for this situation (because of the linear process assumption) we investigate the possible utility of the tapered estimates proposed in exploratory analysis to detect the persistence in some crude approximations to the volatility (like the squares and absolute value of the levels) without the need to model the short run dependence. The above models are strictly stationary for any $0 \leqslant d \leqslant 1$, but a further difficulty is that when $\omega > 0$ the squared process has a drift term and so it is non covariance stationary. We hope that with enough tapering (large p) we can alleviate the effect of this possible drift, which is a smooth function of time t and could be well approximated by polynomials of t.

We do this for two data sets corresponding to the returns (defined as the increment of the logarithm) of the exchange rates of the French franc and the deusch mark against the US dollar, using 2000 daily observations running from November 1972 to January 1981. The plots of the relevant series are given in Figure 3 and the results are given in Figure 4. We employ bandwidth numbers $m = 15, 18, \ldots, 100$ and tapers with $p = 1, 2, 3$. We plot all the estimates obtained in this way, using the squares and the absolute value of the returns series.

The main conclusions we can draw are as follows. Estimates with $p = 1$ usually give a lower range of values than those with higher values of p. In all cases, when we take m too big, the estimates produce much lower values of $\hat{d}$

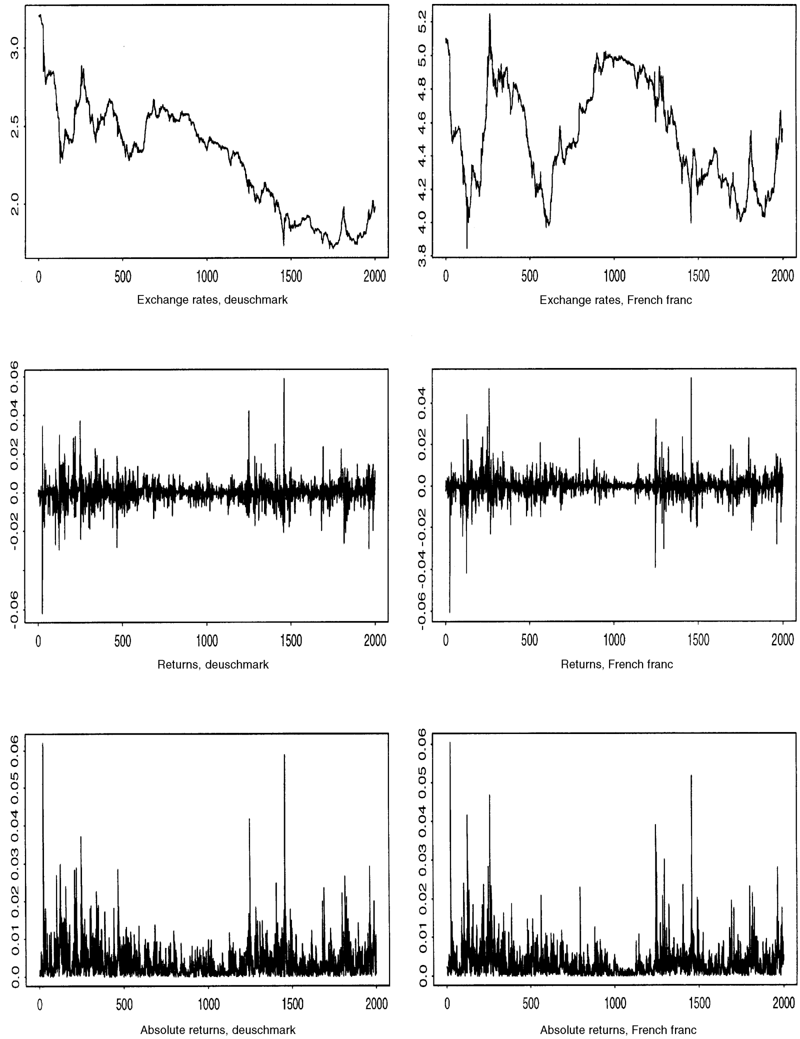

FIGURE 3. Exchange rates, returns and absolute returns for the deutschmark and French franc against the US dollar, November 1973 to January 1981.

as a consequence of moving away from the origin, where we would not expect model (1) to hold. For the significant range of values of m the estimates with $p = 2$ and 3 are almost always very close, indicating perhaps that with $p = 1$ we cannot estimate high values of d appropriately. For the Frech franc the

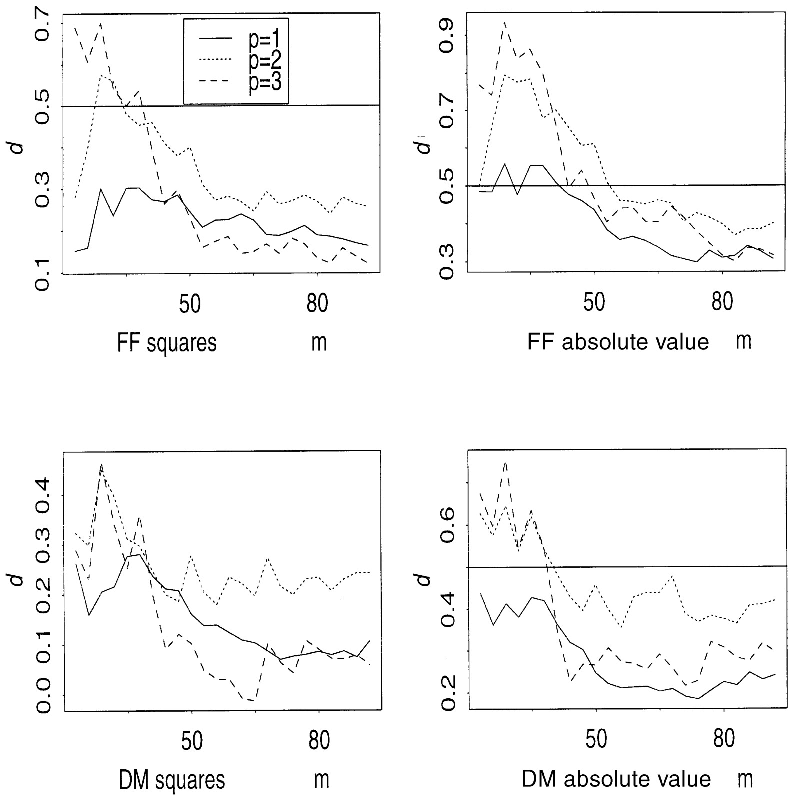

FIGURE 4. Gaussian semiparametric estimates of the persistence for the French franc and the deutschmark.

persistence in volatility is in general higher than for the mark, with values of d up to 0.9 with the absolute value for the franc and only 0.7 for the mark. This agrees with the findings of the previous authors, who reported values for the deutschmark between 0.6 and 0.8 depending on the parametric model assumed for the short run dynamics of the volatility.

8. DISCUSSION

In this paper and in Velasco (1997a) we have shown that the semiparametric model (1) is valid for estimating the memory d of possibly non stationary time series. If the observed process is non stationary $f(\lambda)$ is no longer a spectral density but is the limit of the expectation of the (tapered) periodogram and therefore can be estimated non parametrically. Both the log periodogram and the Gaussian semiparametric estimates compare the non parametric estimate of $f(\lambda)$ given by the periodogram at the relevant frequencies with model (1) and obtain the best estimate of d under different criteria. For this, the integrability or not of the function f around the origin does not matter, but only the accuracy with which we can estimate it by means of the periodogram ordinates. Of course, the steeper and more non integrable f is, the more complicated this approximation will be, but the error can be controlled if enough tapering is applied.

The same principle will undoubtedly work for full parametric models of functions f corresponding to non stationary observations if tapered observations are used. Then, simultaneous estimation of d and the other short run memory parameters is possible without *a priori* assumptions about the degree of (possible) non stationarity of the observed sequence.

Nevertheless this approach will surely break down if we try to estimate the integral below $f(\lambda)$, $\int_0^{\alpha} f(\lambda) d\lambda$ for any $\alpha > 0$, instead of the function f itself, since this integral diverges for $d \geqslant \frac{1}{2}$. This problem arises for the semipara metric estimate of d considered by Robinson (1994a) and Lobato and Robinson (1996), based precisely on the estimation of the cumulative spectral distribution function. Simulations with this estimator $\overline{d}$ always result in estimates of d constrained to $\overline{d} < \frac{1}{2}$, for any $d \geqslant \frac{1}{2}$ and any order of data tapering.

A further approach to deal with long memory, non stationarity and polynomial trends could be the use of wavelets and there are several recent references which deal with the estimation of d and related topics for fractional white noise inference using wavelets (e.g. Jensen, 1995; McCoy and Walden, 1996; and the references therein). Based on the wavelet decomposition of the variance at different scales, a variety of estimates of d are proposed, some close to the log periodogram estimate and others related to Gaussian maximum likelihood, always using the information at all possible scales, being mainly then of full parametric nature. The lack of rigorous asymptotic theory for such estimates in a general case is related to some possible bias problems if the spectral density is not proportional to λ^{-2d} for all frequencies. Furthermore, the assumption of covariance stationarity of the filtered series makes it difficult to predict how these procedures will deal with non stationary observations.

APPENDIX A: PROOFS

PROOF OF THEOREM 2. We repeat the steps of the proof of Theorem 1 in Robinson (1995b), with the same definitions and with the notation in terms of $d \quad H \quad \frac{1}{2}$, readjusting accordingly the set of admissible values $[\nabla_1, \nabla_2]$. More details can be found in that reference or in the proof of Theorem 5. We will concentrate mainly on the asymptotics when $d_0 \geqslant \frac{1}{2}$, since the case $d_0 \in (\ \frac{1}{2}, \frac{1}{2})$ is covered in Robinson's paper.

As in Robinson's proof we define $\nabla \quad \nabla_1$ when $d_0 < \frac{1}{2} + \nabla_1$ and $d_0 \quad \frac{1}{2} < \nabla \leqslant d_0$ otherwise. Then define $\Theta_1 \quad \{d: \nabla \leqslant d \leqslant \nabla_2\}$ and $\Theta_2 \quad \{d: \nabla_1 \leqslant d < \nabla\}$, possibly empty. We take up the proof after expression (3.12) in that reference. Given that now we can consider values of d arbitrarily close to 1, we obtain that for $r \quad 1, 2, \ldots, m$

$$\sup_{\Theta_1} \left| \left(1 + \frac{1}{r}\right)^{2(d-d_0)} \quad 1 \right| \leqslant \frac{12}{r}$$

so the bound is of the same order of magnitude as in the (exclusively) stationary case.

When the observed time series is stationary $d_0 < \frac{1}{2}$, all Robinson's results apply, even if $\nabla_2 \geqslant \frac{1}{2}$. The differences arise when we have to consider the periodogram $I_j \quad I(\lambda_j)$ and $d_0 \geqslant \frac{1}{2}$. When $d_0 < \frac{1}{2}$, we can use expression (3.14) in Robinson's paper,

$$\frac{I_j}{g_j} \quad 1 \quad \left(1 \quad \frac{g_j}{f_j}\right) \frac{I_j}{g_j} + \frac{1}{f_j} (I_j \quad |\alpha_j|^2 I_{\epsilon j}) + (2\pi I_{\epsilon j} \quad 1)$$

where $I_{\epsilon j} \quad I_\epsilon(\lambda_j)$ is the periodogram of $\{\epsilon_t\}_1^n$, $f_j \quad f(\lambda_j)$, $\alpha_j \quad \alpha(\lambda_j)$ and $g_j \quad G\lambda_j^{-2d_0}$. However, when $\frac{1}{2} \leqslant d_0 < 1$ we have to consider the additional transfer function of the linear filter of first differences before writing down the previous decomposition in terms of the sequence ϵ_t:

$$\frac{I_j}{g_j} \quad 1 \quad \left(1 \quad \frac{g_j}{f_j}\right) \frac{I_j}{g_j} + \frac{1}{f_j} \{I_j \quad |1 \quad \exp(i\lambda_j)|^{-2} |\alpha_j|^2 I_{\epsilon j}\} + (2\pi I_{\epsilon j} \quad 1).$$

Now, from Theorem 1 (see also Theorem 1 in Hurvich and Ray, 1995), $d_0 \geqslant \frac{1}{2}$, for n sufficiently large,

$$E \left| \frac{I_j}{g_j} \right| \leqslant C \qquad j \quad 1, \ldots, m \tag{A1}$$

for a generic positive finite constant C, in a similar way to when $d_0 < \frac{1}{2}$.

Next, paralleling expression (3.17) in Robinson (1995b) for the stationary situation,

$$\begin{aligned}
E|I_j \quad & |1 \quad \exp(i\lambda_j)|^{-2} |\alpha_j|^2 I_{\epsilon j}| \\
& \leqslant E[|w_j \quad \{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j w_{\epsilon j}\| w_j + \{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j w_{\epsilon j}|] \\
& \leqslant [EI_j \quad \{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j E w_{\epsilon j} \overline{w}_j \quad \overline{\{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j} E \overline{w}_{\epsilon j} w_j \\
& \quad + |\{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j|^2 EI_{\epsilon j}]^{1/2} \\
& \quad \times [EI_j + \{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j E w_{\epsilon j} \overline{w}_j + \overline{\{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j} E \overline{w}_{\epsilon j} w_j \\
& \quad + |\{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j|^2 EI_{\epsilon j}]^{1/2}
\end{aligned} \tag{A2}$$

denoting by $w_{j\epsilon} \quad w_\epsilon(\lambda_j)$ the Fourier transform of ϵ_t. Then, from the proof of part (a) of Theorem 1 (see Velasco, 1997a) we can obtain, for $\frac{1}{2} \leqslant d_0 < 1$,

$$EI_j \quad f_j\{1 + \mathrm{O}(j^{2(d_0-1)} \log j)\}$$

$$E w_j \overline{w}_{\epsilon j} \quad \frac{\{1 \quad \exp(i\lambda_j)\}^{-1} \alpha_j}{2\pi} + \mathrm{O}(j^{2(d_0-1)} \lambda_j^{-d} \log j)$$

$$EI_{\epsilon j} \quad \frac{1}{2\pi} + \mathrm{O}(j^{2(d_0-1)} \log j)$$

uniformly in $j \quad 1, \ldots, m$. Thus (A2) is $\mathrm{O}\{j^{d_0-1} (\log j)^{1/2}\}$, and following with Robinson's proof, when $d_0 \geqslant \frac{1}{2}$

$$E\left(\sum_{1}^{m-1}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\left|\sum_{1}^{r}\frac{1}{f_j}[I_j \quad |\{1 \quad \exp(i\lambda_j)\}^{-1}\alpha_j|^2 I_{\epsilon j}]\right|\right)$$

$$\leqslant C\sum_{1}^{m}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\sum_{1}^{r}\{j^{d_0-1}(\log j)^{1/2}\}$$

$$\leqslant Cm^{2(d_0-\nabla)-1}\sum_{1}^{m} r^{2(\nabla-d_0)-1+d_0}(\log r)^{1/2}$$

$$\mathrm{O}\{m^{2(d_0-\nabla)-1}+m^{d_0-1}(\log m)^{3/2}\} \quad \mathrm{o}(1)$$

where the last line follows from separate consideration of the cases $2(\nabla \quad d_0) \quad 1+d_0 < \quad 1$ and $2(\nabla \quad d_0) \quad 1+d_0 \geqslant \quad 1$. Also we can check, using the same techniques, that, as $n\to\infty$, for arbitrarily small η, $\frac{1}{2}\leqslant d_0<1$,

$$\left|\frac{1}{m}\sum_{1}^{m}\left(\frac{I_j}{g_j} \quad 1\right)\right| \quad \mathrm{O_P}\left\{\eta+\frac{1}{m}\sum_{1}^{m} j^{d_0-1}(\log m)^{1/2}\right\}+\mathrm{o_P}(1) \quad \mathrm{o_P}(1).$$

Using Robinson's definitions the next point that deserves attention when $d_o \geqslant \frac{1}{2}+\nabla_1$ is

$$\left|\frac{1}{m}\sum_{1}^{m}(a_j \quad 1)\left(1 \quad \frac{g_j}{f_j}\right)\frac{I_j}{g_j}\right| \quad \mathrm{O_P}\left\{\frac{\eta}{m}\sum_{1}^{m}(a_j+1)\right\} \quad \mathrm{O_P}(\eta)$$

with (A1).

Observe that after Equation (3.22) in Robinson (1995b) we need to choose in fact $\nabla < d_0 \quad \frac{1}{2}+1/(4e)$ without loss of generality. Because of this modification, we have to proceed in a different way to bound the next expression, for $\frac{1}{2}\leqslant d_0<1$:

$$\left|\frac{1}{m}\sum_{1}^{m}\frac{a_j \quad 1}{f_j}[I_j \quad |\{1 \quad \exp(i\lambda_j)\}^{-1}\alpha_j|^2 I_{\epsilon j}]\right| \tag{A3}$$

$$\mathrm{O_P}\left\{\frac{1}{m}\sum_{1}^{m}(a_j+1)j^{d_0-1}(\log m)^{1/2}\right\}$$

$$\mathrm{O_P}\left\{\frac{1}{m}\sum_{1}^{m} a_j j^{d_0-1}(\log m)^{1/2}+\frac{1}{m}\sum_{1}^{m} j^{d_0-1}(\log m)^{1/2}\right\}. \tag{A4}$$

Next, since $p \quad \exp(m^{-1}\sum_1^m \log j)\sim m/e$,

$$\sum_{1}^{p} a_j j^{d_0-1} \quad p^{2(d_0-\nabla)}\sum_{1}^{p} j^{2(\nabla-d_0)+d_0-1} \quad \mathrm{O}(m^{d_0})$$

if $2\nabla \quad d_0>0$, and $\mathrm{O}(m^{2(d_0-\nabla)}\log m)$ if $2\nabla \quad d_0\leqslant 0$. Then, using $\sum_p^m a_j \quad \mathrm{O}(m)$ and $\sup_{j>p} j^{d_0-1} \quad \mathrm{O}(p^{d_0-1}) \quad \mathrm{O}(m^{d_0-1})$, we obtain that (A4) is

$$\mathrm{O_P}\{m^{-1}(m^{d_0}+m^{2(d_0-\nabla)})(\log m)^{3/2}\} \quad \mathrm{o_P}(1)$$

with $d_0<1$ and $d_0 \quad \frac{1}{2}<\nabla$, and the proof is completed.

Proof of Theorem 3. Again we retrace the steps in the proof of Theorem 2 in Robinson (1995b). The main step here is to obtain the equivalent to expression (4.7) in that proof bounding in probability the quantity

$$\sum_{1}^{r}\left(\frac{I_j}{g_j} \quad 2\pi I_{\epsilon j}\right)$$

for the general case $d_0 \in (\ \frac{1}{2}, \frac{3}{4})$. We shall see that the bounds for the case $d_o \geqslant \frac{1}{2}$ are weaker in general than for the stationary case, so these will be the leading terms in the bounds.

First, we need the quantity (cf. Equation (4.7) in Robinson (1995b)), for $0 < \delta < \frac{1}{2}$,

$$\sum_{r \ \ 1}^{m}\left(\frac{r}{m}\right)^{1-2\delta}\frac{1}{r^2}\left|\sum_{1}^{r}\left(\frac{I_j}{g_j} \quad 1\right)\right| + \frac{1}{m}\left|\sum_{1}^{m}\left(\frac{I_j}{g_j} \quad 1\right)\right| \tag{A5}$$

to be $o_P\{(\log m)^{-6}\}$. From Lemma 1, the second term in (A5) is, for $d_0 \geqslant \frac{1}{2}$,

$$O_P\{m^{4(d_0-1)/(5-4d_0)}(\log m)^{2/(5-4d_0)} + m^{\beta}n^{-\beta}$$
$$+ m^{2(d_0-1)}\log m + n^{-1/2}m^{(d_0-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_0-1}(\log m)^{1/2}\}$$
$$o_P\{(\log m)^{-6}\}$$

if $d_0 < 1$, with (4), and the first term is in order of probability

$$m^{2\delta-1}\sum_{r \ \ 1}^{m} r^{-1-2\delta}\{r^{1/(5-4d_0)}(\log r)^{2/(5-4d_0)} + r^{\beta+1}n^{-\beta}$$
$$+ r^{2d_0-1}\log r + n^{-1/2}r^{(1+d_0)/2}(\log n)^{5/4} + n^{-1/4}r^{d_0}(\log r)^{1/2}\}$$
$$O(m^{2\delta-1}[1 + m^{1-2\delta}\{m^{4(d_0-1)/(5-4d_0)}(\log m)^{2/(5-4d_0)} + m^{\beta}n^{-\beta}$$
$$+ m^{2(d_0-1)}\log m + n^{-1/2}m^{(d_0-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_0-1}(\log m)^{1/2}\}])$$
$$o_P\{(\log m)^{-6}\}.$$

From Lemma 1, we can see also with $\hat{F}_k(d) \quad m^{-1}\sum_1^m(\log j)^k\lambda_j^{2d}I_j$,

$$\left|\hat{F}_k(d_0) \quad G_0\frac{1}{m}\sum_{1}^{m}(\log j)^k\right|$$
$$O_P[\{m^{4(d_0-1)/(5-4d_0)} + m^{2(d_0-1)} + n^{-1/2}m^{(d_0-1)/2}(\log n)^{5/4} + n^{-1/4}m^{d_0-1}\}(\log m)^2]$$
$$o_P(1)$$

if $\frac{1}{2} \leqslant d_0 < 1$. Next, the error in probability after expression (4.11) in Robinson's proof is now with Lemma 1

$$O_P[\{m^{(4d_0-3)/(10-8d_0)}(\log m)^{3/2} + m^{\beta+1/2}n^{-\beta}$$
$$+ m^{2d_0-3/2}\log m + n^{-1/2}m^{d_0/2}(\log n)^{5/4} + n^{-1/4}m^{d_0-1/2}(\log m)^{1/2}\}\log m]$$
$$o_P(1)$$

if $\frac{1}{2} \leqslant d_0 < \frac{3}{4}$, using (4). This completes the proof using the same central limit theorem.

PROOF OF THEOREM 5. We repeat the steps of the proof of Theorem 1 in Robinson (1995b), with the same definitions and with the notation in terms of $d \quad H \quad \frac{1}{2}$, readjusting accordingly the set of admissible values $[\nabla_1, \nabla_2]$.

For $\frac{1}{2} > \delta > 0$ let $N_\delta \quad \{d\colon |d \quad d_0| < \delta\}$ and $\overline{N}_\delta \quad (\ \infty, \infty) \quad N_\delta$. Then, for $S_p(d) \quad R_p(d) \quad R_p(d_0)$,

$$
\begin{aligned}
P(|\hat{d} \quad d_0| \geqslant \delta) \quad & P(\hat{d} \in \overline{N}_\delta \cap \Theta) \\
& P\{\inf_{\overline{N}_\delta \cap \Theta} R_p(d) \leqslant \inf_{N_\delta \cap \Theta} R_p(d)\} \\
& \leqslant P\{\inf_{\overline{N}_\delta \cap \Theta} S_p(d) \leqslant 0\}
\end{aligned}
$$

because $d_0 \in N_\delta \cap \Theta$. As in Robinson's proof, we define $\nabla \quad \nabla_1$ when $d_0 < \frac{1}{2} + \nabla_1$ and $d_0 \quad \frac{1}{2} < \nabla \leqslant d_0$ otherwise. Then $\Theta_1 \quad \{d: \nabla \leqslant d \leqslant \nabla_2\}$ and $\Theta_2 \quad \{d: \nabla_1 \leqslant d < \nabla\}$, possibly empty. It follows that

$$
P(|\hat{d} \quad d_0| \geqslant \delta) \leqslant P\{\inf_{\overline{N}_\delta \cap \Theta_1} S_p(d) \leqslant 0\} + P\{\inf_{\Theta_2} S_p(d) \leqslant 0\}. \tag{A6}
$$

The sets Θ_1 and Θ_2 are treated separately because of the non uniform behaviour of $R_p(d)$ around $d \quad d_0 \quad \frac{1}{2}$. The first probability on the right of (A6) is bounded by

$$
P\{\sup_{\Theta_1} |T_p(d)| \geqslant \inf_{\overline{N}_\delta \cap \Theta_1} U_p(d)\} \tag{A7}
$$

where

$$
\begin{aligned}
T_p(d) \quad & \log\left\{\frac{\hat{G}(d)}{G_0}\right\} \quad \log\left\{\frac{\overline{G}(d_0)}{G(d)}\right\} \quad \log\left\{\frac{2(d \quad d_0) + 1}{m^{2(d-d_0)}} \frac{p}{m} \sum_j^m j^{2(d-d_0)}\right\} \\
& + 2(d \quad d_0)\left\{\frac{p}{m} \sum_j^m \log j \quad (\log m \quad 1)\right\} \\
U_p(d) \quad & 2(d \quad d_0) \quad \log\{2(d \quad d_0) + 1\} \\
G_p(d) \quad & G_0 \frac{p}{m} \sum_j^m \lambda_j^{2(d-d_0)}
\end{aligned}
$$

so that $S_p(d) \quad U_p(d) \quad T_p(d)$. As in Robinson (1995b),

$$
\inf_{\overline{N}_\delta \cap \Theta_1} U_p(d) > \tfrac{1}{2}\delta^2 \tag{A8}
$$

and $\sup_{\overline{N}_\delta \cap \Theta_1} |T_p(d)| \to_{\mathrm{p}} 0$ if

$$
\sup_{\Theta_1} \left| \frac{\hat{G}_p(d) \quad G_p(d)}{G_p(d)} \right| \tag{A9}
$$

is $o_P(1)$, while

$$
\sup_{\Theta_1} \left| \frac{p\{2(d \quad d_0) + 1\}}{m} \sum_j^m \left(\frac{j}{m}\right)^{2(d-d_0)} \quad 1 \right| \tag{A10}
$$

and

$$
\left| \frac{p}{m} \sum_j^m \log m \quad (\log m \quad 1) \right| \tag{A11}
$$

are both o(1).

From Lemmas 4 and 5 below, (A10) and (A11) are $O(m^{-2(\nabla - d_0) - 1}) \quad o(1)$ and $O(\log m / m) \quad o(1)$ as $m \to \infty$, respectively. We write

$$\frac{\hat{G}_p(d) \quad G_p(d)}{G_p(d)} \quad \frac{A_p(d)}{B_p(d)}$$

where

$$A_p(d) \quad \frac{p\{2(d \quad d_0)+1\}}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_0)}\left(\frac{I_j}{g_j} \quad 1\right)$$

$$B_p(d) \quad \frac{p\{2(d \quad d_0)+1\}}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_0)}$$

for $g_j \quad G_0\lambda_j^{-2d_0}$. Now

$$\inf_{\Theta_1} B_p(d) \geqslant 1 \quad \sup_{\Theta_1}\left|\frac{p\{2(d \quad d_0)+1\}}{m}\sum_j^m \left(\frac{j}{m}\right)^{2(d-d_0)} \quad 1\right| \geqslant \frac{1}{2} \tag{A12}$$

for all sufficiently large m, by Lemma 4. By summation by parts

$$|A_p(d)| \leqslant \frac{3p}{m}\left|\sum_r^{m-p}\left\{\left(\frac{r}{m}\right)^{2(d-d_0)} \quad \left(\frac{r+p}{m}\right)^{2(d-d_0)}\right\}\sum_j^r\left(\frac{I_j}{g_j} \quad 1\right)\right| + \frac{3p}{m}\left|\sum_j^m\left(\frac{I_j}{g_j} \quad 1\right)\right|. \tag{A13}$$

Because $|(1+1/r)^{2(d-d_0)} \quad 1| \leqslant C_{\nabla_2,p}/r$ on Θ_1 when $r>0$ where $C_{\nabla_2,p}$ is a constant depending on ∇_2 and p such that

$$C_{\nabla_2,p} \leqslant (2\nabla_2+1)\left(\frac{p+1}{p}\right)^{2\nabla_2}$$

the first term on the right of (A13) has supremum on Θ_1 bounded by

$$3C_{\nabla_2,p}\,p\sup_{\Theta_1}\sum_r^{m-p}\left(\frac{r}{m}\right)^{2(d-d_0)+1}\frac{1}{r^2}\left|\sum_j^r\left(\frac{I_j}{g_j} \quad 1\right)\right|$$
$$\leqslant 3C_{\nabla_2,p}\,p\sum_r^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\left|\sum_j^r\left(\frac{I_j}{g_j} \quad 1\right)\right| \tag{A14}$$

where the inequality is due to $0<2(\nabla \quad d_0)+1 \leqslant 2(d \quad d_0)+1$ on Θ_1.

Now we have to consider the periodogram $I_j^{\mathrm{T}} \quad I_p^{\mathrm{T}}(\lambda_j)$ in the decomposition

$$\frac{I_j^{\mathrm{T}}}{g_j} \quad 1 \quad \left(1 \quad \frac{g_j}{f_j}\right)\frac{I_j^{\mathrm{T}}}{g_j}+\frac{1}{f_j}\{I_j^{\mathrm{T}} \quad |1 \quad \exp(i\lambda_j)|^{-2s}|\alpha_j|^2 I_{\epsilon j}^{\mathrm{T}}\}+(2\pi I_{\epsilon j}^{\mathrm{T}} \quad 1). \tag{A15}$$

For any $\eta>0$, Assumptions 7 and 8 imply that n can be chosen such that

$$\left|1 \quad \frac{g_j}{f_j}\right| \leqslant \eta \qquad j \quad 1, \ldots, m. \tag{A16}$$

Now, from the proof of Theorem 4 in Velasco (1997a), for n sufficiently large,

$$E\left|\frac{I_j^{\mathrm{T}}}{g_j}\right| \leqslant C \qquad j \quad 1, \ldots, m \tag{A17}$$

for a generic positive finite constant C, in a similar way to when $d_0 \in (\quad \frac{1}{2}, \frac{1}{2})$. Thus

$$E\left\{\sum_{r}^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\left|\sum_{j}^{r}\left(1 \quad \frac{g_j}{f_j}\right)\frac{I_j^{\mathrm{T}}}{g_j}\right|\right\}\leqslant\frac{C\eta}{2(\nabla \quad d_0)+1}.$$

Next, generalizing expression (A2),

$$\begin{aligned}
&E|I_j^{\mathrm{T}} \quad |1 \quad \exp(i\lambda_j)|^{-2s}|\alpha_j|^2 I_{\epsilon j}^{\mathrm{T}}|\\
&\qquad\leqslant E[w_j^{\mathrm{T}} \quad \{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j w_{\epsilon j}^{\mathrm{T}}\|w_j^{\mathrm{T}}+\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j w_{\epsilon j}^{\mathrm{T}}]\\
&\qquad\leqslant[EI_j^{\mathrm{T}} \quad \{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j E w_{\epsilon j}^{\mathrm{T}}\overline{w}_j^{\mathrm{T}} \quad \overline{\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j}E\overline{w}_{\epsilon j}^{\mathrm{T}}w_j^{\mathrm{T}}\\
&\qquad\quad+|\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j|^2 EI_{\epsilon j}^{\mathrm{T}}]^{1/2}\\
&\qquad\quad\times[EI_j^{\mathrm{T}}+\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j E w_{\epsilon j}^{\mathrm{T}}\overline{w}_j^{\mathrm{T}}+\overline{\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j}E\overline{w}_{\epsilon j}^{\mathrm{T}}w_j^{\mathrm{T}}\\
&\qquad\quad+|\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j|^2 EI_{\epsilon j}^{\mathrm{T}}]^{1/2} \qquad\qquad (\mathrm{A18})
\end{aligned}$$

denoting by $w_{j\epsilon}^{\mathrm{T}} \quad w_{\epsilon}^{\mathrm{T}}(\lambda_j)$ the (tapered) Fourier transform of ϵ_t. Then, from the proof of part (a) in Theorem 4 (see Velasco, 1997a) we obtain that, as $n\to\infty$,

$$EI_j^{\mathrm{T}} \quad f_j\{1+\mathrm{O}(j^{-1}+j^{2(d_0-p)}\log j)\}$$

$$Ew_j^{\mathrm{T}}\overline{w}_{\epsilon j}^{\mathrm{T}} \quad \frac{\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j}{2\pi}+\mathrm{O}(j^{-1}\lambda_j^{-d_0}+j^{2(d_0-p)}\lambda_j^{-d_0}\log j)$$

$$EI_{\epsilon j}^{\mathrm{T}} \quad \frac{1}{2\pi}+\mathrm{O}(j^{-1}+j^{2(d_0-p)}\log j)$$

uniformly in $j \quad p, 2p, \ldots, m$. Thus (A18) is $\mathrm{O}(f_j[j^{-1/2}+j^{d_0-p}\{\log(j+1)\}^{1/2}])$, and following Robinson's proof

$$\begin{aligned}
&E\left(\sum_{1}^{m-p}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\left|\sum_{1}^{r}\frac{1}{f_j}[I_j \quad |\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j|^2 I_{\epsilon j}]\right|\right)\\
&\qquad\leqslant C\sum_{1}^{m}\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\sum_{1}^{r}\{j^{-1/2}+j^{d_0-p}(\log j)^{1/2}\}\\
&\qquad\leqslant Cm^{2(d_0-\nabla)-1}\sum_{1}^{m}\{r^{2(\nabla-d_0)-1/2}+r^{2(\nabla-d_0)-p+d_0}(\log j)^{1/2}\}\\
&\qquad\quad \mathrm{O}\{m^{2(d_0-\nabla)-1}+m^{-1/2}\log m+m^{d_0-p}(\log m)^{3/2}\} \quad \mathrm{o}(1)
\end{aligned}$$

with $d_0<p$, where the last line follows from separate consideration of the cases $2(\nabla \quad d_0) \quad \frac{1}{2}< \quad 1$ and $2(\nabla \quad d_0) \quad \frac{1}{2}\geqslant \quad 1$, and $2(\nabla \quad d_0) \quad p+d_0< \quad 1$ and $2(\nabla \quad d_0) \quad p+d_0\geqslant \quad 1$.

To deal with the final contribution to (A15) we need to consider the variance of $\sum_j^r(2\pi I_{\epsilon j}^{\mathrm{T}} \quad 1)$, since it has zero mean. The variance and covariances of $I_{\epsilon j}^{\mathrm{T}}$ have two components. The first is due to the fourth cumulant

$$\left(\sum_{1}^{n}h_t^2\right)^{-2}\int_{[-\pi,\pi]^3}D_p^{\mathrm{T}}(\omega_1+\lambda_j)D_p^{\mathrm{T}}(\omega_2 \quad \lambda_k)D_p^{\mathrm{T}}(\omega_3 \quad \lambda_j)D_p^{\mathrm{T}}\left(\lambda_k \quad \sum_{1,2,3}\omega_i\right)$$
$$\times f_\epsilon^{(4)}(\omega_1,\omega_2,\omega_3)d\omega$$

which is of order n^{-1}, given the boundedness of $f_\epsilon^{(4)}$ and the properties of D_p^{T},

$\int_{-\pi}^{\pi}|D_p^{\mathrm{T}}(\lambda)|d\lambda = O(1)$ and $\sup_\lambda|D_p^{\mathrm{T}}(\lambda)| = O(n)$, all n and $p>1$. The second component is due to the second moments. The variance of $I_{\epsilon j}^{\mathrm{T}}$ is then $O(1)$ and for the covariance between $I_{\epsilon j}^{\mathrm{T}}$ and $I_{\epsilon k}^{\mathrm{T}}$, $k \neq j$, in addition to the $O(n^{-1})$ fourth cumulant term, we have to consider the convolutions

$$\left(\sum_1^n h_t^2\right)^{-2}\int_{-\pi}^{\pi} D_p^{\mathrm{T}}(\lambda+\lambda_j)D_p^{\mathrm{T}}(\lambda-\lambda_k)d\lambda\int_{-\pi}^{\pi} D_p^{\mathrm{T}}(\lambda-\lambda_j)D_p^{\mathrm{T}}(\lambda+\lambda_k)d\lambda$$

and

$$\left(\sum_1^n h_t^2\right)^{-2}\int_{-\pi}^{\pi} D_p^{\mathrm{T}}(\lambda+\lambda_j)D_p^{\mathrm{T}}(\lambda+\lambda_k)d\lambda\int_{-\pi}^{\pi} D_p^{\mathrm{T}}(\lambda-\lambda_j)D_p^{\mathrm{T}}(\lambda-\lambda_k)d\lambda$$

since $f_\epsilon(\lambda)$ is constant. These terms, from Lemmas 1 and 2 of Velasco (1997a), are $\mathrm{O}(|j-k|^{-2p})$ and $\mathrm{O}(|j+k|^{-2p})$, respectively, for $j,\ k>0$. Thus

$$\begin{aligned}\operatorname{var}\left\{\sum_j^r(2\pi I_{\epsilon j}^{\mathrm{T}}-1)\right\} &= \sum_j^r \operatorname{var}(2\pi I_{\epsilon j}^{\mathrm{T}})+\sum_{j\neq}^r\sum_k^r \operatorname{cov}(2\pi I_{\epsilon j}^{\mathrm{T}},\ 2\pi I_{\epsilon k}^{\mathrm{T}})\\ &= \mathrm{O}(r)+\mathrm{O}\left\{\sum_{j\neq}^r\sum_k^r(|j-k|^{-2p}+|j+k|^{-2p}+n^{-1})\right\} = \mathrm{O}(r).\end{aligned}$$

We have obtained that $\sum_j^r(2\pi I_{\epsilon j}^{\mathrm{T}}-1) = \mathrm{O_P}(r^{1/2})$ and so

$$\begin{aligned}\sum_1^m\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}\frac{1}{r^2}\left|\sum_1^r(2\pi I_{\epsilon j}-1)\right| &= \mathrm{O_P}\left\{\sum_1^m\left(\frac{r}{m}\right)^{2(\nabla-d_0)+1}r^{-3/2}\right\}\\ &= \mathrm{O_P}\left(m^{2(d_0-\nabla)-1}\sum_1^m r^{2(\nabla-d_0)-1/2}\right)\\ &= \mathrm{O_P}(m^{-1/2}+m^{2(d_0-\nabla)-1}\log m) = \mathrm{o_P}(1)\end{aligned}$$

as $n\to\infty$ because $2(d_0-\nabla)<1$.

Also we can check, using the same techniques, that, as $n\to\infty$, for arbitrarily small η, since $d_0<p$,

$$\left|\frac{1}{m}\sum_1^m\left(\frac{I_j}{g_j}-1\right)\right| = \mathrm{O_P}\left\{\eta+\frac{1}{m}\sum_1^m j^{d_0-p}(\log m)^{1/2}\right\}+\mathrm{o_P}(1) = \mathrm{o_P}(1).$$

Thus, as $n\to\infty$, $\sup_{\Theta_1}|A_p(d)|\to_{\mathrm{P}} 0$ and, with (A9) and (A12), $\sup_{\Theta_1}|\hat{G}_p(d)/G_p(d)-1|\to_{\mathrm{P}} 0$. In view of (A8) it follows that (A7) $\to 0$ as $n\to\infty$.

When $d_0\geqslant\frac{1}{2}+\nabla_1$ we have to consider the second probability on the right of (A6). Set $q = q_m = \exp(pm^{-1}\sum_j^m\log j)$ and $S_p(d) = \log\{\hat{D}_p(d)/\hat{D}_p(d_0)\}$, where

$$\hat{D}_p(d) = \frac{p}{m}\sum_j^m\left(\frac{j}{q}\right)^{2(d-d_0)}j^{2d_0}I_j^{\mathrm{T}}.$$

Because $1\leqslant q\leqslant m$ and $\inf_{\Theta_2}(j/q)^{2(d-d_0)}\geqslant(j/q)^{2(\nabla-d_0)}$ for $1\leqslant j\leqslant q$, while $\inf_{\Theta_2}(j/q)^{2(d-d_0)}\geqslant(j/q)^{2(\nabla_1-d_0)}$ for $q<j\leqslant m$, it follows that

$$\inf_{\Theta_2}\hat{D}_p(d)\geqslant\frac{p}{m}\sum_j^m a_j j^{2d_0}I_j^{\mathrm{T}}$$

where

$$a_j \quad \begin{cases} \left(\dfrac{j}{q}\right)^{2(\nabla - d_0)} & 1 \leqslant j \leqslant q \\ \left(\dfrac{j}{q}\right)^{2(\nabla_1 - d_0)} & q < j \leqslant m. \end{cases}$$

Thus

$$P\{\inf_{\Theta_2} S_p(d) \leqslant 0\} \leqslant P\left\{\frac{p}{m}\sum_j^m (a_j \quad 1) j^{2d_0} I_j \leqslant 0\right\}.$$

As $m \to \infty$, $q \sim \exp(\log m \quad 1) \quad m/e$ and

$$\sum_{1 \leqslant j \leqslant q} a_j \sim p^{-1} q^{2(d_0 - \nabla)} \int_0^q x^{2(\nabla - d_0)} dx \quad \frac{q/p}{2(\nabla \quad d_0) + 1} \sim \frac{(m/e)/p}{2(\nabla \quad d_0) + 1}. \tag{A19}$$

It follows that

$$\frac{p}{m}\sum_j^m (a_j \quad 1) \geqslant \frac{p}{m}\sum_{1 \leqslant j \leqslant q} a_j \quad 1 \sim \frac{1}{e\{2(\nabla \quad d_0) + 1\}} \quad 1 \qquad \text{as } m \to 0.$$

Choose $\nabla < d_0 \quad \frac{1}{2} + 1/(4e)$, which we may do with no loss of generality. Then for all sufficiently large m, $(p/m)\sum_j^m (a_j \quad 1) \geqslant 1$ and thus (A6) is bounded by

$$P\left\{\left|\frac{p}{m}\sum_j^m (a_j \quad 1)\left(\frac{I_j}{g_j} \quad 1\right)\right| \geqslant 1\right\}.$$

Now apply (A15) again and first note from (A16) and (A17) that

$$\left|\frac{p}{m}\sum_j^m (a_j \quad 1)\left(1 \quad \frac{g_j}{f_j}\right)\frac{I_j}{g_j}\right| \quad \mathrm{O_P}\left\{\frac{\eta}{m}\sum_j^m (a_j + 1)\right\} \quad \mathrm{O_P}(\eta)$$

and

$$\sum_{q < j < m} a_j \sim p^{-1} q^{2(d_0 - \nabla_1)} \int_q^m x^{2(\nabla_1 - d_0)} dx \quad \mathrm{O}(m)$$

and

$$\sum_j^m a_j^2 \quad \mathrm{O}(m^{4(d_0 - \nabla)} + m \log m).$$

Observe that after Equation (2.9) in Robinson (1995b) we need to choose in fact $\nabla < d_0 \quad \frac{1}{2} + 1/(4e)$ and not $\nabla < d_0 \quad \frac{1}{2} + e/4$, without loss of generality. Because of this modification, we have to proceed in a different way to bound the expression

$$\left|\frac{p}{m}\sum_1^m \frac{a_j \quad 1}{f_j}[I_j \quad |\{1 \quad \exp(i\lambda_j)\}^{-s}\alpha_j|^2 I_{\epsilon j}]\right| \tag{A20}$$

$$\mathrm{O_P}\left[\frac{1}{m}\sum_1^m (a_j + 1)\{j^{-1/2} + j^{d_0 - p}(\log m)^{1/2}\}\right]$$

$$\mathrm{O_P}\left\{\frac{1}{m}\sum_1^m a_j(j^{-1/2} + j^{d_0 - p})(\log m)^{1/2} + m^{-1/2} + \frac{1}{m}\sum_1^m j^{d_0 - p}(\log m)^{1/2}\right\}. \tag{A21}$$

Next, since $q \sim m/(ep)$,

$$\sum_{1}^{q} a_j j^{d_0 - p} \quad q^{2(d_0-\nabla)} \sum_{1}^{q} j^{2(\nabla - d_0) + d_0 - p} \quad \mathrm{O}(m^{d_0 - p+1})$$

if $2(\nabla \quad d_0) + d_0 \quad p > 0$, and $\mathrm{O}(m^{2(d_0-\nabla)} \log m)$ if $2(\nabla \quad d_0) + d_0 \quad p \leqslant 0$. Also

$$\sum_{1}^{q} a_j j^{-1/2} \quad q^{2(d_0-\nabla)} \sum_{1}^{q} j^{2(\nabla - d_0)-1/2} \quad \mathrm{O}(m^{1/2})$$

if $2(\nabla \quad d_0) \quad 1/2 > 0$, and $\mathrm{O}(m^{2(d_0-\nabla)} \log m)$ if $2(\nabla \quad d_0) \quad 1/2 \leqslant 0$.

Then, using $\sum_q^m a_j \quad \mathrm{O}(m)$ and $\sup_{j>q} j^{d_0-p} \quad \mathrm{O}(q^{d_0-p}) \quad \mathrm{O}(m^{d_0-p})$, we obtain that (A21) is

$$\mathrm{O_P}\{m^{-1/2} + m^{-1}(m^{d_0-p+1} + m^{1/2} + m^{2(d_0-\nabla)})(\log m)^{3/2}\} \quad \mathrm{o_P}(1)$$

with $d_0 < p$ and $d_0 \quad \frac{1}{2} < \nabla$.

Finally, using Theorem 4 and proceeding as before,

$$\mathrm{var}\left\{\frac{p}{m}\sum_{j}^{m}(a_j \quad 1)(2\pi I_{\epsilon j}^{\mathrm{T}} \quad 1)\right\}$$

$$\frac{p^2}{m^2}\sum_{j}^{m}(a_j \quad 1)^2 \mathrm{var}(2\pi I_{\epsilon j}^{\mathrm{T}}) + \frac{p^2}{m^2}\sum_{j/}^{m}\sum_{k}^{m}(a_j \quad 1)(a_k \quad 1)\mathrm{cov}(2\pi I_{\epsilon j}^{\mathrm{T}}, 2\pi I_{\epsilon k}^{\mathrm{T}})$$

$$\mathrm{O}\left\{m^{-2}\sum_{j}^{m}(a_j \quad 1)^2 \sum_{k/j}^{m}(|j \quad k|^{-2p} + |j+k|^{-2p})\right\}$$

$$\mathrm{O}\left\{m^{-2}\left(m + \sum_{j}^{m} a_j^2\right)\right\}$$

$$\mathrm{O}\{m^{-2}(m \log m + m^{4(d_0-\nabla)})\}$$

$$\mathrm{O}(m^{-1}\log m + m^{2\{2(d_0-\nabla)-1\}}) \quad \mathrm{o}(1)$$

and the proof is completed.

Proof of Theorem 6. We can adapt all the steps in the proof of Theorem 2 in Robinson (1995b) to the situation for $p > 1$ as we have done in the proof of Theorem 5. This amounts basically to redefinition of the sums to frequencies $\lambda_p, \lambda_{2p}, \ldots, \lambda_m$ only.

The main step here is to bound in probability the quantity (cf. Robinson, 1995b, Equation (4.7)), for $0 < \delta < \frac{1}{2}$,

$$A\sum_{r \quad 1}^{m}\left(\frac{r}{m}\right)^{1-2\delta}\frac{1}{r^2}\left|\sum_{1}^{r}\left(\frac{I_j}{g_j} \quad 1\right)\right| + B\frac{1}{m}\left|\sum_{1}^{m}\left(\frac{I_j}{g_j} \quad 1\right)\right| \tag{A22}$$

to be $\mathrm{o_P}\{(\log m)^{-6}\}$, where A and B are two finite constants depending on p and ∇_2 (see Equation (A14) above).

Now, using the same procedure as in the proof of Theorem 5 (cf. Robinson's Equation (3.17) and the following text), using $\beta > 1$ with $r \leqslant m$,

$$\sum_{p}^{r}\left(\frac{I_j}{g_j} \quad 2\pi I_{\epsilon j}\right) \quad \mathrm{O_P}(r^{1-\beta/2} + \log r + r^{d_0-p+1}(\log r)^{1/2} + r^{\beta+1}n^{-\beta}) \tag{A23}$$

where the term $\log r$ shows up when $\beta \quad 2$ (or when $d_0 \quad p \quad 1$), and the term $r^{\beta+1}n^{-\beta}$ is exactly the same as in Robinson's expression (4.8) (see also the equation after (4.25)). Note that in this case we have followed a much more direct approach than Robinson's (1995b) proof, using a stronger assumption on the smoothness of the function f, namely $\beta > 1$. This is in part for convenience and in part because the correlation between adjacent tapered periodogram ordinates invalidates the approach using second moments of the periodogram as in Robinson (1995b, p. 1648) and used above when $p \quad 1$ and $d_0 < \frac{3}{4}$. A similar approach was used in Velasco (1997a) to analyse the log periodogram ordinate for non Gaussian stationary observations. Note that this procedure is only valid if we use a tapered periodogram with $p \geqslant 2$ but not otherwise: we use the lower bias of tapering, avoiding the increment of correlation.

The bound in probability at the end of page 1643 in Robinson (1995b) is now, using (A22) and (A23) as $n \to \infty$,

$$\mathrm{O_P}[\{m^{-1/2}\log m + m^{d_0-p}(\log m)^{3/2} + m^{\beta}n^{-\beta}\}(\log m)^2] \quad \mathrm{o_P}(1).$$

From here we can reach the same limit as in expression (4.10), and the equivalent to expression (4.11) in Robinson's paper is now

$$\left(2\left(\frac{m}{p}\right)^{-1/2}\sum_{p}^{m}\nu_j(2\pi I_{\epsilon j} \quad 1)\right.$$
$$\left. + \mathrm{O_P}[\{m^{(1-\beta)/2}(\log m)^2 + m^{d_0-p+1/2}(\log m)^2 + m^{\beta+1/2}n^{-\beta}\}\log m]\right)\{1+\mathrm{o_P}(1)\}$$

where $\nu_j \quad \log j \quad (p/m)\sum_p^m \log j$ satisfies $\sum_p^m \nu_j \quad 0$, which, from the assumptions of the theorem, is

$$\left\{2\left(\frac{m}{p}\right)^{-1/2}\sum_{p}^{m}\nu_j(2\pi I_{\epsilon j} \quad 1) + \mathrm{o_P}(1)\right\}\{1+\mathrm{o_P}(1)\}.$$

Using Lemma 6 we can obtain the asymptotic distribution of $(m/p)^{-1/2}\sum_p^m \nu_j(2\pi I_{\epsilon j} \quad 1)$ and the theorem is proved.

APPENDIX B: TECHNICAL LEMMAS

LEMMA 1. *Under the Assumptions of Theorem 3,* $d_0 \in [\frac{1}{2}, 1)$,

$$\sum_{1}^{r}\left(\frac{I_j}{g_j} \quad 2\pi I_{\epsilon j}\right) \quad \mathrm{O_P}\{r^{1/(5-4d_0)}(\log r)^{2/(5-4d_0)} + r^{\beta+1}n^{-\beta}$$
$$+ r^{2d_0-1}\log r + n^{-1/2}r^{(1+d_0)/2}(\log n)^{5/4} + n^{-1/4}r^{d_0}(\log r)^{1/2}\}$$

PROOF. We only consider the case $d_0 \geqslant \frac{1}{2}$, since the stationary situation follows as in the proof of Theorem 2 in Robinson (1995b), with stronger results. Choosing an integer $1 < l < r$, for $d_0 \in [\frac{1}{2}, 1)$ from (A1) and with $E(2\pi I_{\epsilon j}) \quad 1$,

$$E\left|\sum_{1}^{l}\left(\frac{I_j}{g_j} \quad 2\pi I_{\epsilon j}\right)\right| \quad \mathrm{O}(l)$$

and also from (A1) and Assumption 2

$$E\left|\sum_{l+1}^{r}\left(\frac{I_j}{g_j}\quad\frac{I_j}{f_j}\right)\right| \leqslant C\sum_{l+1}^{r}\left|1\quad\frac{g_j}{f_j}\right|\quad \mathrm{O}\left(\frac{r^{\beta+1}}{n^{\beta}}\right).$$

Next, we consider

$$E\left[\left\{\sum_{l+1}^{r}\left(\frac{I_j}{f_j}\quad 2\pi I_{\epsilon j}\right)\right\}^2\right]\quad (2\pi)^2(a+b)$$

with the same definitions as in Robinson (1995b, p. 1648). Further, if we split the terms $a \quad a_1+a_2$ and $b \quad b_1+b_2$ corresponding to second and fourth cumulants, we find that when $d_0\in[\frac{1}{2}, 1)$ with Theorem 1

$$a_1\quad \mathrm{O}\left(\sum_{l+1}^{r} j^{2(d_0-1)}\log j\right)\quad \mathrm{O}\{r^{2d_0-1}(\log r)^2\}$$

and

$$b_1\quad \mathrm{O}\left[\sum_{j\quad l+1}^{r}\sum_{k>j}^{r}\{(jk)^{2(d_0-1)}(\log k)^2+(j^{2(d_0-1)}\log k)^2\}\right]$$

$$\mathrm{O}\left\{(\log r)^2\sum_{k\quad l+2}^{r}\sum_{j\quad l+1}^{k-1} j^{4(d_0-1)}\right\}$$

$$\mathrm{O}\{rl^{4d_0-3}(\log r)^2\}$$

since we will only use $r\quad \mathrm{O}(n)$ at most. Choosing $l\sim r^{1/(5-4d_0)}(\log r)^{2/(5-4d_0)}$ this gives the first term of this order in the lemma, since $(a_1)^{1/2}$ is of smaller order of magnitude.

When $d_0\geqslant\frac{1}{2}$ we obtain the same expressions for a_2 and b_2 as in Robinson (1995b), and substituting $\alpha(\lambda)$ by $\{1\quad \exp(i\lambda)\}^{-1}\alpha(\lambda)$ and α_j by $\{1\quad \exp(i\lambda_j)\}^{-1}\alpha_j$ and defining

$$P_j\quad \int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda)\{1\quad \exp(i\lambda_j)\}}{\{1\quad \exp(i\lambda)\}\alpha_j}\quad 1\right|^2 K(\lambda\quad \lambda_j)d\lambda$$

where $K(\lambda)\quad (2\pi n)^{-1}\sin(n\lambda/2)\sin(\lambda/2)$ is the Fejér kernel, the same bounds hold here. However, the bound for the second type of summand considered by Robinson in b_2, $\mathrm{O}(P_jP_k^{1/2})$, is improved in Lemma 2 to $\mathrm{O}\{n^{-1}P_k^{1/2}(\log n)^2\}$. This allows consideration of values of the parameter $d_0<\frac{3}{4}$, which otherwise would be restricted to $d_0<\frac{2}{3}$. For a_2 we can still use the bound given by Robinson.

Then applying Lemma 3, with Lemma 2,

$$a_2\quad \mathrm{O}\left[\sum_{j\quad 1}^{r}\left\{\frac{(\log j)^2}{j^{4(1-d_0)}}+\frac{(\log j)^{3/2}}{j^{3(1-d_0)}}+\frac{n^{-1/2}\log j}{j^{2(1-d_0)}}\right\}\right]$$

$$\mathrm{O}\{r^{3d_0-2}(\log r)^{3/2}+n^{-1/2}r^{2d_0-1}\log j+(\log r)^3\}$$

$$b_2\quad \mathrm{O}\left[\sum_{j\quad 1}^{r}\sum_{k>j}^{r}\left\{\frac{(\log r)^2}{(jk)^{2(1-d_0)}}+\frac{(\log k)^{1/2}(\log n)^2}{nk^{1-d_0}}+\frac{n^{-1/2}\log r}{(jk)^{1-d_0}}\right\}\right]$$

$$\mathrm{O}\{r^{2(2d_0-1)}(\log r)^2+n^{-1}r^{1+d_0}(\log n)^{5/2}+n^{-1/2}r^{2d_0}\log r+(\log r)^4\}$$

and the lemma follows.

LEMMA 2. *Under Assumptions 1 and 5,* $d_0 \in [\frac{1}{2}, 1)$,

$$\frac{1}{(2\pi n)^2}\int_{\Pi^3}\left(\frac{\alpha(\lambda+\mu+\zeta)\{1 \quad \exp(i\lambda_j)\}}{[1 \quad \exp\{i(\lambda+\mu+\zeta\}]\alpha_j} \quad 1\right)\left[\frac{\alpha(\quad \mu)\{1 \quad \exp(\quad i\lambda_j)\}}{\{1 \quad \exp(\quad i\mu)\}\overline{\alpha}_j} \quad 1\right]$$

$$\times\left[\frac{\alpha(\quad \zeta)\{1 \quad \exp(\quad i\lambda_k)\}}{\{1 \quad \exp(\quad i\zeta)\}\overline{\alpha}_k} \quad 1\right]\times E_{jk}(\lambda, \mu, \zeta)d\lambda d\mu d\zeta$$

$$\mathrm{O}\{n^{-1}k^{d_0-1}(\log n)^2(\log k)^{1/2}\} \qquad \text{(B1)}$$

where

$$E_{jk}(\lambda, \mu, \zeta) \quad D(\lambda_j \quad \lambda \quad \nu \quad \zeta)D(\lambda_k+\lambda)D(\mu \quad \lambda_j)D(\zeta \quad \lambda_k)$$

and $D(\lambda) \quad \sum_t \exp(i\lambda t)$ *is the Dirichlet kernel.*

PROOF. Making a change of variable and using the periodicity of D, (B1) is

$$\frac{1}{(2\pi n)^2}\int_{\Pi^3}\left[\frac{\alpha(\omega)\{1 \quad \exp(i\lambda_j)\}}{\{1 \quad \exp(i\omega)\}\alpha_j} \quad 1\right]\left[\frac{\alpha(\quad \mu)\{1 \quad \exp(\quad i\lambda_j)\}}{\{1 \quad \exp(\quad i\mu)\}\overline{\alpha}_j} \quad 1\right]$$

$$\times\left[\frac{\alpha(\quad \zeta)\{1 \quad \exp(\quad i\lambda_k)\}}{\{1 \quad \exp(\quad i\zeta)\}\overline{\alpha}_k} \quad 1\right]$$

$$\times D(\lambda_j \quad \omega)D(\lambda_k+\omega \quad \mu \quad \zeta)D(\mu \quad \lambda_j)D(\zeta \quad \lambda_k)d\omega d\mu d\zeta$$

and this is less in absolute value than

$$\frac{1}{2\pi n}P_k^{1/2}\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)\{1 \quad \exp(i\lambda_j)\}}{\{1 \quad \exp(i\omega)\}\alpha_j} \quad 1\right||D(\lambda_j \quad \omega)|d\omega\int_{-\pi}^{\pi}\left|\frac{\alpha(\quad \mu)\{1 \quad \exp(\quad i\lambda_j)\}}{\{1 \quad \exp(\quad i\mu)\}\overline{\alpha}_j} \quad 1\right|$$

$$\times|D(\mu \quad \lambda_k)|d\mu.$$

Now using the bound for P_k in Lemma 3 and

$$\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)\{1 \quad \exp(i\lambda_j)\}}{\{1 \quad \exp(i\omega)\}\alpha_j} \quad 1\right|D(\lambda_j \quad \omega)d\omega \quad \mathrm{O}(\log n) \qquad \text{(B2)}$$

the lemma follows. To prove (B2) we consider now

$$\int_{-\pi}^{\pi}\left|\frac{\alpha(\omega)\{1 \quad \exp(i\lambda_j)\}}{\{1 \quad \exp(i\omega)\}\alpha_j} \quad 1\right||D(\lambda_j \quad \omega)|d\omega$$

$$\leqslant\left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|^{-1}\int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}} \quad \frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right||D(\omega)|d\omega$$

and the following intervals of integration:

$$\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right|\leqslant\left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|^{-1}\sup_{-\lambda_j/2\leqslant\omega\leqslant\lambda_j/2}\left|\frac{d}{d\omega}\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}}\right|\int_{-\lambda_j/2}^{\lambda_j/2}|\omega||D(\omega)|d\omega$$

$$\mathrm{O}(\lambda_j^d\lambda_j^{-d_0-1}\lambda_j) \quad \mathrm{O}(1).$$

Next

$$\left|\int_{\lambda_j/2}^{3\lambda_j/2}\right| \leqslant \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|^{-1} \int_{\lambda_j/2}^{3\lambda_j/2}\left[\left|\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}}\right| \quad \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|\right]|D(\omega)|d\omega$$

$$\mathrm{O}\left(\sup_{\lambda_j/2\leqslant\omega\leqslant 3\lambda_j/2}|D(\omega)|\left[\lambda_j^{d}\int_{\lambda_j/2}^{3\lambda_j/2}\left|\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}}\right|d\omega + \int_{\lambda_j/2}^{3\lambda_j/2} d\omega\right]\right)$$

$$\mathrm{O}\left\{\lambda_j^{-1}\left(\lambda_j^{d_0}\int_0^{\lambda_j}\omega^{-d_0}d\omega + \lambda_j\right)\right\} \qquad \mathrm{O}(1)$$

since $d_0 < 1$ (note that $|\alpha(\lambda)\{1 \quad \exp(i\lambda)\}^{-1}| \quad \{2\pi f(\lambda)\}^{1/2}$ is integrable because $d_0 < 1$). Then, choosing $\epsilon > 0$, fixed, as small as we want, such that Assumption 1 holds for $|\lambda| < \epsilon$, as in the proof of Theorem 2 of Robinson (1995a),

$$\left|\int_{-\epsilon}^{-\lambda_j/2}\right| \leqslant \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|^{-1} \sup_{-\epsilon\leqslant\omega\leqslant-\lambda_j/2}\left[\left|\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}}\right| \quad \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|\right]$$

$$\times\int_{-\pi}^{\pi}|D(\omega)|d\omega \qquad \mathrm{O}(\log n)$$

$$\left|\int_{3\lambda_j/2}^{\epsilon}\right| \leqslant \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|^{-1} \sup_{3\lambda_j/2\leqslant\omega\leqslant\epsilon}\left[\left|\frac{\alpha(\lambda_j \quad \omega)}{1 \quad \exp\{i(\lambda_j \quad \omega)\}}\right| \quad \left|\frac{\alpha(\lambda_j)}{1 \quad \exp(i\lambda_j)}\right|\right]$$

$$\times\int_{-\pi}^{\pi}|D(\omega)|d\omega \qquad \mathrm{O}(\log n)$$

and the same bound holds for the remaining intervals of integration

LEMMA 3. *Under Assumptions 1 and 5, with* $d_0 \in [\frac{1}{2}, 1)$,

$$P_j \qquad \int_{-\pi}^{\pi}\left|\frac{\alpha(\lambda)\{1 \quad \exp(i\lambda_j)\}}{\{1 \quad \exp(i\lambda)\}\alpha_j} \quad 1\right|^2 K(\lambda \quad \lambda_j)d\lambda \qquad \mathrm{O}(j^{2(d_0-1)}\log j).$$

PROOF. This is Lemma 3 of Robinson (1995b) generalized to cover the non stationary situation $d_0 \in [\frac{1}{2}, 1)$ and follows considering the same intervals of integration, where for the interval $[\quad \lambda_j/2, \lambda_j/2]$ we can adapt the proofs of Theorem 1 or Theorem 6 $(p \quad 1)$ in Velasco (1997a), since $f(\lambda)$ is not integrable at the origin, to obtain an $\mathrm{O}(j^{2(d_0-1)}\log j)$ contribution.

LEMMA 4. *For* $p \quad 1, 2, \ldots, \varepsilon \in (0, 1]$ *and* $C \in (\varepsilon, \infty)$*, as* $m \to \infty$,

$$\sup_{\varepsilon\leqslant\gamma\leqslant C}\left|\frac{\gamma p}{m}\sum_{j \quad p,2p,\ldots}^{m}\left(\frac{j}{m}\right)^{\gamma-1} \quad 1\right| \quad \mathrm{O}\left(\frac{1}{m^{\varepsilon}}\right).$$

PROOF. As in Lemma 1 of Robinson (1995b), $\int_0^a x^{\gamma-1}dx \quad a^\gamma/\gamma$ for $\gamma > 0$,

$$\left|\frac{\gamma p}{m}\sum_{j\quad p,2p,\ldots}^{m}\left(\frac{j}{m}\right)^{\gamma-1}\quad 1\right| \leqslant \gamma\int_0^{p/m}\left\{\left(\frac{p}{m}\right)^{\gamma-1}\quad x^{\gamma-1}\right\}dx$$

$$+\gamma\sum_{j\quad 2p,3p,\ldots}^{m}\int_{(j-p)/m}^{j/m}\left\{\left(\frac{j}{m}\right)^{\gamma-1}\quad x^{\gamma-1}\right\}dx$$

$$\leqslant \frac{\gamma}{(m/p)^{\gamma}}+\frac{1}{(m/p)^{\gamma}}+\frac{\gamma|\gamma\quad 1|}{(m/p)^2}\sum_{j\quad p}^{m}\left(\frac{j}{m}\right)^{\gamma-2}$$

by the mean value theorem. The last term is $O(\gamma^2 m^{-1})$ for $\gamma > 1$, zero for $\gamma\quad 1$ and $O(m^{-\gamma})$ for $0 < \gamma < 1$.

LEMMA 5. *For all* $m \geqslant 2p,\ p\quad 2, 3, \ldots,$

$$\left|\frac{p}{m}\sum_{j\quad p,2p,\ldots}^{m}\log j\quad \log m+1\right| \leqslant \frac{2p\log p\quad p+1+\log(m\quad p)}{m}. \tag{B3}$$

PROOF. Because $\int_0^m \log x dx\quad m(\log m\quad 1)$, the left hand side of (B3) is

$$\left|\frac{p\log p}{m}+\frac{1}{m}\int_0^p \log x dx\quad \frac{1}{m}\sum_{j\quad 2p,3p,\ldots}^{m}\int_{j-p}^{j}\log\left(\frac{j}{x}\right)dx\right|$$

$$\leqslant \frac{p(2\log p\quad 1)}{m}+\frac{1}{m}\sum_{j\quad p,2p,\ldots}^{m-p}\frac{1}{j}$$

$$\leqslant \frac{p(2\log p\quad 1)}{m}+\frac{1+\log(m\quad p)}{m}$$

$$\frac{2p\log p\quad p+1+\log(m\quad p)}{m}.$$

LEMMA 6. *If the sequence* $\{h_j\}$ *is a data taper of order p as defined previously, and the random variables* $\{\epsilon_j\}$ *satisfy Assumption 6, with* $\nu_j\quad \log j\quad (p/m)\sum_{j\quad p}^{m}\log j$,

$$Z_n\quad \left(\frac{m}{p}\right)^{-1/2}\sum_{p}^{m}\nu_j\{2\pi I_\epsilon^{\mathrm{T}}(\lambda_j)\quad 1\} \to_{\mathrm{D}} N(0, \Phi)$$

where Φ *is given in (10).*

PROOF. We will follow Robinson (1995b, pp. 1644 47), adapting his non tapered proof to the tapered case. We have that $Z_n\quad 2\sum_{t\quad 1}^{n} z_t$ and

$$z_t\quad h_t\epsilon_t\sum_{s\quad 1}^{t-1}h_s\epsilon_s c_{t-s}$$

$$c_s\quad 2\left(\sum_{r\quad 1}^{n}h_r^2\right)^{-1}\left(\frac{m}{p}\right)^{-1/2}\nu_j\cos(s\lambda_j)$$

remembering that $\sum_{r\quad 1}^{n}h_r^2 \sim bn$. Now the z_t form a zero mean martingale difference array, and from a standard central limit theorem we can deduce that $\sum z_t$ tends to an $N(0, \Phi)$ random variable in distribution if

$$\sum_{t\quad 1}^{n} E(z_t^2 | F_{t-1}) \qquad \Phi \to_{\mathrm{p}} 0 \tag{B4}$$

$$\sum_{t\quad 1}^{n} E\{z_t^2 I(|z_t| > \rho)\} \to 0 \qquad \text{for all } \rho > 0. \tag{B5}$$

Not the left hand side of (B4) is

$$\left(\sum_{t\quad 1}^{n} h_t^2 \sum_{s\quad 1}^{t-1} h_s^2 \epsilon_s^2 c_{t-s}^2 \qquad \Phi \right) + \sum_{t\quad 1}^{n} h_t^2 \sum_{s\quad 1}^{t-1} \sum_{r/s}^{t-1} h_s \epsilon_s h_r \epsilon_r c_{t-s} c_{t-r}. \tag{B6}$$

The term in parentheses is

$$\left\{ \sum_{t\quad 1}^{n-1} h_t^2 (\epsilon_t^2 \qquad 1) \sum_{s\quad 1}^{n-t} h_{s+t}^2 c_s^2 \right\} + \left\{ \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 c_s^2 \qquad \Phi \right\}. \tag{B7}$$

Now

$$\begin{aligned} \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 c_s^2 \qquad & \frac{4p}{m(\sum_r h_r^2)^2} \sum_{j\quad p}^{m} \sum_{k\quad p}^{m} \nu_j \nu_k \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 \cos(s\lambda_j) \cos(s\lambda_k) \\ & \frac{4p}{m(\sum_r h_r^2)^2} \sum_{j\quad p}^{m} \nu_j^2 \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 \cos^2(s\lambda_j) \\ & + \frac{2p}{m(\sum_r h_r^2)^2} \sum_{j\quad p}^{m} \sum_{k/j}^{m} \nu_j \nu_k \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 [\cos\{s(\lambda_j \qquad \lambda_k)\} \\ & + \cos\{s(\lambda_j + \lambda_k)\}]. \end{aligned}$$

Next, using part (A) of Lemma 7, for n large enough,

$$\frac{4p}{m} \left(\sum_r h_r^2 \right)^{-2} \sum_{j\quad p}^{m} \nu_j^2 \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 \cos^2(s\lambda_j) \qquad \frac{p}{m} \sum_{j\quad p}^{m} \nu_j^2 + \mathrm{O}\{m^{-1}(\log m)^2 + n^{-1}\} \tag{B8}$$

and using part (B) of Lemma 7

$$\begin{aligned} & \frac{2p}{m} \left(\sum_r h_r^2 \right)^{-2} \sum_{j\quad p}^{m} \sum_{k/j}^{m} \nu_j \nu_k \sum_{t\quad 1}^{n-1} h_t^2 \sum_{s\quad 1}^{n-t} h_{s+t}^2 [\cos\{s(\lambda_j \qquad \lambda_k)\} + \cos\{s(\lambda_j + \lambda_k)\}] \\ & \qquad \frac{p}{m} \left(\sum_r h_r^2 \right)^{-2} \sum_{j}^{m} \sum_{k/j}^{m} \nu_j \nu_k \left(\left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \qquad \lambda_k)\} \right]^2 + \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j + \lambda_k)\} \right]^2 \right) \\ & \qquad + \mathrm{O}\{mn^{-1}(\log m)^2\}. \end{aligned} \tag{B9}$$

Noting that, for $1 \leqslant j \leqslant n/2$,

$$\left(\sum_{1}^{n} h_t^2 \right)^{-1} \sum_{1}^{n} h_t^2 \cos(t\lambda_j) \qquad \mathrm{O}(j^{-p}) \tag{B10}$$

(see, for example, Lemmas 1 and 2 in Velasco, 1997a), then

$$\frac{p}{m}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \sum_{k/j}^{m} \nu_j \nu_k \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j + \lambda_k)\}\right]^2 \quad \mathrm{O}\left\{ m^{-1} \sum_{j\;\;p}^{m} \sum_{k\;\;p}^{m} \nu_j \nu_k (j+k)^{-2p} \right\}$$

$$\mathrm{O}\left\{ m^{-1} (\log m)^2 \sum_{j\;\;p}^{m} \sum_{k\;\;p}^{m} j^{-p} k^{-p} \right\}$$

$$\mathrm{O}\{ m^{-1} (\log m)^2 \}$$

and the second term in brackets in (B9) can be neglected. For the other term in (B9) we can write, including simultaneously the first component of the right hand side of (B8), for $0 \leqslant \eta(n) \leqslant m$,

$$\frac{p}{m}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \sum_{k\;\;p}^{m} \nu_j \nu_k \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \quad \lambda_k)\}\right]^2$$

$$\frac{p}{m}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \sum_{k:|j-k|\leqslant\eta} \nu_j \nu_k \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \quad \lambda_k)\}\right]^2$$

$$+\frac{p}{m}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \sum_{k:|j-k|>\eta} \nu_j \nu_k \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \quad \lambda_k)\}\right]^2$$

$$\frac{p}{m}\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \nu_j^2 \sum_{k:|j-k|\leqslant\eta} \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \quad \lambda_k)\}\right]^2$$

$$+\mathrm{O}\left(m^{-1} \left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{j\;\;p}^{m} \sum_{k:|j-k|\leqslant\eta} |\nu_j| \sup_{|j-k|\leqslant\eta} |\nu_j \quad \nu_k| \left[\sum_{1}^{n} h_t^2 \cos\{t(\lambda_j \quad \lambda_k)\}\right]^2 \right)$$

$$+\mathrm{O}\left\{ m^{-1} (\log m)^2 \sum_{j\;\;p}^{m} \sum_{k:|j-k|>\eta} |j \quad k|^{-2p} \right\}$$

and this is

$$\frac{p}{m}\sum_{j\;\;p}^{m} \nu_j^2 \left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{k\quad 0,p,2p,\ldots}^{n-p} \left\{\sum_{1}^{n} h_t^2 \cos(t\lambda_k)\right\}^2$$

$$+\mathrm{O}\left(m^{-1} \sum_{j\;\;p}^{m} \nu_j^2 \sum_{k>\eta}^{n} k^{-2p} \right) + \mathrm{O}\left(m^{-1} \log m \sum_{j\;\;p}^{m} \frac{\eta}{j} \sum_{k:|j-k|\leqslant\eta} |j \quad k|^{-2p} \right)$$

$$+\mathrm{O}\{ \eta^{1-2p} (\log m)^2 \}$$

which using

$$\frac{p}{m}\sum_{p}^{m} \nu_j^2 \quad 1 + \mathrm{O}\left\{\frac{(\log m)^2}{m}\right\}$$

is

$$\left(\sum_{1}^{n} h_t^2\right)^{-2} \sum_{k\;\;0,p,2p,\ldots}^{n-p} \left\{\sum_{1}^{n} h_t^2 \cos(t\lambda_k)\right\}^2 + \mathrm{O}(\eta^{1-2p}) + \mathrm{O}\left\{\frac{\eta}{m}(\log m)^2\right\}$$

$$+\ \mathrm{O}\{\eta^{1-2p}(\log m)^2\} + \mathrm{o}(1) \quad\ \Phi + \mathrm{o}(1).$$

The errors are o(1) on choosing, for example, $\eta \sim m^{1/2}$, and $\Phi < \infty$ exists due to (B10). Then the second term in (B7) is o(1) as $n \to \infty$. The first component of (B7) has zero mean and variance

$$\mathrm{O}\left\{\sum_{t\;\;1}^{n-1} h_t^2 \left(\sum_{s\;\;1}^{n-t} h_{s+t}^2 c_s^2\right)^2\right\}.$$

Now using the same bounds for $|c_s|$ as in Robinson (1995b) and noting that $\sup_t |h_t| \leqslant 1$, we obtain that this is $\mathrm{O}\{(\log n)^4/n\}$, so (B7) is $\mathrm{o_P}(1)$. The second component of (B6) has zero mean and variance

$$2\sum_{t\;\;2}^{n} h_t^2 \sum_{u\;\;2}^{n} h_u^2 \sum_{s}^{\min(t-1,u-1)} \sum_{r/s} h_s^2 h_r^2 c_{t-r} c_{t-s} c_{u-r} c_{u-s}$$

$$2\sum_{t\;\;2}^{n} h_t^4 \sum_{s} \sum_{r/s} h_s^2 h_r^2 c_{t-r}^2 c_{t-s}^2 + 4\sum_{t\;\;3}^{n} h_t^2 \sum_{u\;\;2}^{t-1} h_u^2 \sum_{s}^{u-1} \sum_{r/s}^{u-1} h_s^2 h_r^2 c_{t-r} c_{t-s} c_{u-r} c_{u-s}$$

because the weights $\{h_t\}$ are symmetric around $\lfloor n/2 \rfloor$. As in Robinson's paper, the first term on the right is $\mathrm{O}\{(\log m)^4/n\}$ and the second has absolute value bounded by

$$4\sum_{t\;\;3}^{n} \sum_{u\;\;2}^{t-1} \left(\sum_{s}^{u-1} c_{t-r}^2 \sum_{r/s}^{u-1} c_{u-r}^2\right) \leqslant 4\left(\sum_{1}^{n} c_t^2\right)\left(\sum_{t\;\;3}^{n} \sum_{u\;\;2}^{t-1} \sum_{r\;\;t-u+1}^{t-1} c_r^2\right)$$

since $\sup_t |h_t| \leqslant 1$, and using the same arguments as in that reference this is $\mathrm{O}\{(\log m)^4/m^{1/3}\}$ and thus we have verified (B4). To prove (B5) we also check the sufficient condition

$$\sum_{1}^{n} E(z^4) \to 0 \qquad \text{as } n \to \infty.$$

The left hand side of this equals

$$\mu_4 \sum_{2}^{n} E\left\{\left(\sum_{1}^{t-1} h_s \epsilon_s c_{t-s}\right)^4\right\}$$

$$\leqslant C\sum_{2}^{n} E\left(\sum_{s\;\;1}^{t-1} \sum_{r\;\;1}^{t-1} \sum_{q\;\;1}^{t-1} \sum_{p\;\;1}^{t-1} h_s h_r h_q h_p \epsilon_s \epsilon_r \epsilon_q \epsilon_p c_{t-s} c_{r-s} c_{q-s} c_{p-s}\right)$$

$$\leqslant C\sum_{1}^{n}\left(\sum_{1}^{n} c_{t-s}^4\right) + C\sum_{1}^{n} \sum_{1}^{t-1} \sum_{1}^{t-1} c_{t-s}^2 c_{t-r}^2$$

$$\mathrm{O}\left\{\frac{(\log m)^4}{n}\right\}$$

using the bound for h_t and the given reference, completing the proof.

LEMMA 7. *If the sequence* $\{h_j\}$ *is a data taper of order* p *as defined previously, for* $0 < |j| < n/2$,

(A)

$$\sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1}^{n-t} h_{s+t}^2 \cos^2(s\lambda_j) \quad \frac{1}{4}\left(\sum_{t\ 1}^{n} h_t^2\right)^2 + \mathrm{O}(n^2 j^{-2p} + n)$$

and, for $0 < |j| < n$,

(B)

$$\sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1}^{n-t} h_{s+t}^2 \cos(s\lambda_j) \quad \frac{1}{2}\left\{\sum_{1}^{n} h_t^2 \cos(t\lambda_j)\right\}^2 + \mathrm{O}(n).$$

PROOF OF (B). We have

$$\begin{aligned} \sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1}^{n-t} h_{s+t}^2 \cos(s\lambda_j) &\quad \sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1-t}^{0} h_{s+t}^2 \cos(s\lambda_j) + \mathrm{O}(n) \\ &\quad \frac{1}{2}\sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1-t}^{n-t} h_{s+t}^2 \cos(s\lambda_j) + \mathrm{O}(n) \\ &\quad \frac{1}{2}\sum_{t\ 1}^{n} h_t^2 \cos(t\lambda_j) \sum_{s\ 1}^{n} h_s^2 \cos(s\lambda_j) + \mathrm{O}(n). \end{aligned}$$

The first two lines follow by symmetry, because $h_t \quad h_{n-t}$ and $\psi_t \quad \phi_{n-t}$, where $\psi_t \quad \sum_{s\ 1}^{n-t} h_{s+t}^2 \cos(s\lambda_j)$ and $\phi_t \quad \sum_{s\ 1-t}^{0} h_{s+t}^2 \cos(s\lambda_j)$, the error terms are due to end effects, and the last step follows because

$$\begin{aligned} \sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1-t}^{n-t} h_{s+t}^2 \cos(s\lambda_j) &\quad \sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1}^{n} h_s^2 \cos\{(s \quad t)\lambda_j\} \\ &\quad \sum_{t\ 1}^{n-1} h_t^2 \sum_{s\ 1}^{n} h_s^2 \{\cos(s\lambda_j)\cos(t\lambda_j) + \sin(s\lambda_j)\sin(t\lambda_j)\} \\ &\quad \sum_{t\ 1}^{n} h_t^2 \sum_{s\ 1}^{n} h_s^2 \cos(s\lambda_j)\cos(t\lambda_j) + \mathrm{O}(n) \end{aligned}$$

since the sine terms cancel out by symmetry again.

PROOF OF (A). Again, by symmetry, changing variable in the sum index, using trigonometric identities and the proof of property (B),

$$
\begin{aligned}
\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n-t} h_{s+t}^2 \cos^2(s\lambda_j) &= \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1-t}^{n-t} h_{s+t}^2 \cos^2(s\lambda_j) + \mathrm{O}(n) \\
&= \frac{1}{2}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \cos^2\{(s-t)\lambda_j\} + \mathrm{O}(n) \\
&= \frac{1}{4}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 [1 - \cos\{2(s-t)\lambda_j\}] + \mathrm{O}(n) \\
&= \frac{1}{4}\sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 - \sum_{t=1}^{n-1} h_t^2 \sum_{s=1}^{n} h_s^2 \cos\{(s-t)\lambda_{2j}\} + \mathrm{O}(n) \\
&= \frac{1}{4}\left(\sum_{t=1}^{n} h_t^2\right)^2 + \mathrm{O}\left[n + \left\{\sum_{s=1}^{n} h_s^2 \cos(s\lambda_{2j})\right\}^2\right]
\end{aligned}
$$

and the lemma follows on using (B10).

ACKNOWLEDGEMENTS

I am grateful to P. M. Robinson, an Associate Editor, and a referee for helpful discussions and suggestions. I also wish to thank J. Arteche and A. Pérez for valuable comments. Part of this research was carried out while the author was at the Departments of Statistics, London School of Economics and University of Oxford. The research was supported by ESRC Grant 235892.

REFERENCES

AGIAKLOGLOU, C., NEWBOLD, P. and WOHAR, M. (1993) Bias in an estimator of the fractional difference parameter. *J. Time Ser. Anal.* 14, 235–46.

ALEKSEEV, V. G. (1996) Jackson and Jackson Vallée Poussin-type kernels and their probability applications. *Theory Probab. Appl.* 41, 137–43.

BAILLIE, R. T., BOLLERSLEV, T. and MIKKELSEN, H. O. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Economet.* 74, 3–30.

BLOOMFIELD, P. (1991) Time series methods. In *Statistical Theory and Modeling* (eds D. V. Hinkley, N. Reid and E. J. Snell). London: Chapman and Hall, pp. 152–76.

BOLLERSLEV, T. and MIKKELSEN, H. O. (1996) Modeling and pricing long memory in stock market volatility. *J. Economet.* 73, 151–84.

DAHLHAUS, R. (1985) Asymptotic normality of spectral estimates. *J. Multivariate Anal.* 16, 412–31.

—— (1988) Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Stat.* 16, 808–41.

FOX, R. and TAQQU, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat.* 14, 517–32.

GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–38.

GIRAITIS, L. and SURGAILIS, D. (1990) A central limit theorem for quandratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Relat. Fields* 86, 87–104.

——, ROBINSON, P. M. and SAMAROV, A. (1995) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* 18, 49–60.

HASSLER, U. (1992) Unit root test: the autoregressive approach in comparison with the periodogram regression. *Stat. Pap.* 34, 67–82.

HURVICH, C. M. and RAY, B. K. (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.* 16, 17–42.

JENSEN, M. J. (1995) Ordinary least squares estimate of the fractional differencing parameter using wavelets as derived from smoothing kernels. Southern Illinois University, Carbondale.

KÜNSCH, H. R. (1987) Statistical aspects of self-similar processes. *Proceedings 1st World Congress of the Bernoulli Society.* (eds Y. A. Prohorov and V. V. Sazonov) Utrecht: VNU Science Press, pp. 67–74.

LOBATO, I. and ROBINSON, P. M. (1996) Averaged periodogram estimation of long memory. *J. Economet.* 73, 303–24.

MCCOY, E. J. and WALDEN, A. T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *J. Comp. Graph. Stat.* 5, 26–56.

PERCIVAL, D. B. and WALDEN, A. T. (1993) *Spectral Analysis for Physical Applications*. Cambridge University Press, Cambridge.

ROBINSON, P. M. (1986) On the errors-in-variables problem for time series. *J. Multivariate Anal.* 19, 240–50.

—— (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Economet.* 47, 67–84.

—— (1994a) Semiparametric analysis of long-memory time series. *Ann. Stat.* 22, 515–39.

—— (1994b) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probab. Theory Relat. Fields* 99, 443–73.

—— (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048–72.

—— (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* 23, 1630–61.

VELASCO, C. (1997a) Non-stationary log-periodogram regression. Forthcoming *J. Economet.*

—— (1997b) Non-Gaussian log-periodogram regression. Preprint.

ZHURBENKO, I. G. (1979) On the efficiency of estimates of a spectral density. *Scand. J. Stat.* 6, 49–56.